**Optimization of Short and Medium-Term Lightning Forecasting through the Integration of Artificial Intelligence and Physical Parameterization Models**

Ikromjon Rakhmonov 1, 2,a), Nurbek Kurbonov1, Li Qingmin 3, Zhang Li 4, Kultan Jaroslav4, Shakhruz Ganiev 7, Khurshid Sattarov 7, Lyazat Sadykova 6

1 Tashkent state technical university named after Islam Karimov, Tashkent, Uzbekistan

2 Karakalpak State University named after Berdakh, Nukus, Uzbekistan

3 North China Electric Power University, Beijing, China

4Shandong University, Jinan, China

5 University of Economics in Bratislava, Bratislava, Slovakia

6 West Kazakhstan Innovative and Technological University, Uralsk, Kazakhstan

7Bukhara State University, Bukhara, Uzbekistan

8Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan

a) Corresponding author: [ilider1987@yandex.com](mailto:ilider1987@yandex.com)

**Abstract.** Lightning constitutes one of the most critical atmospheric threats to large-scale wind energy installations, leading to severe equipment degradation, unexpected outages, and diminished system reliability. Consequently, reliable short- and medium-range lightning prediction is a key prerequisite for implementing proactive protection strategies, real-time monitoring, and fault prevention mechanisms in modern wind farms. Conventional physics-based lightning parameterization schemes embedded within numerical weather prediction frameworks offer strong physical interpretability; however, they frequently struggle to accurately resolve localized convective extremes. In contrast, purely artificial intelligence–based approaches often achieve high short-term predictive accuracy but are limited by reduced robustness, generalization capability, and physical transparency.To overcome these limitations, this study proposes a hybrid AI–physics forecasting framework designed to optimize lightning prediction across short-term (0–2 h) and medium-term (6–24 h) time horizons. The developed methodology combines physically grounded lightning parameterizations with probabilistic machine learning models through optimized data fusion techniques. Meteorological, geographic, and observational datasets are subjected to advanced feature engineering and dimensionality reduction procedures to enhance computational efficiency while preserving physical coherence. Model performance is systematically evaluated against standalone physical and AI-based benchmarks using widely accepted verification metrics, including Probability of Detection, False Alarm Ratio, Critical Success Index, Area Under the Receiver Operating Characteristic Curve, and Brier Score. The obtained results indicate that the hybrid framework delivers consistently improved predictive skill and reliability, increasing detection performance by up to 18% while simultaneously reducing false alarms and forecast uncertainty. Developed within the scope of the project “Advanced Technologies for Lightning Protection-Based Monitoring, Prognosis and Fault Diagnosis in Large-Scale Wind Farms,” the proposed approach offers a scientifically robust and operationally feasible solution for advanced lightning risk assessment in large-scale renewable energy systems.

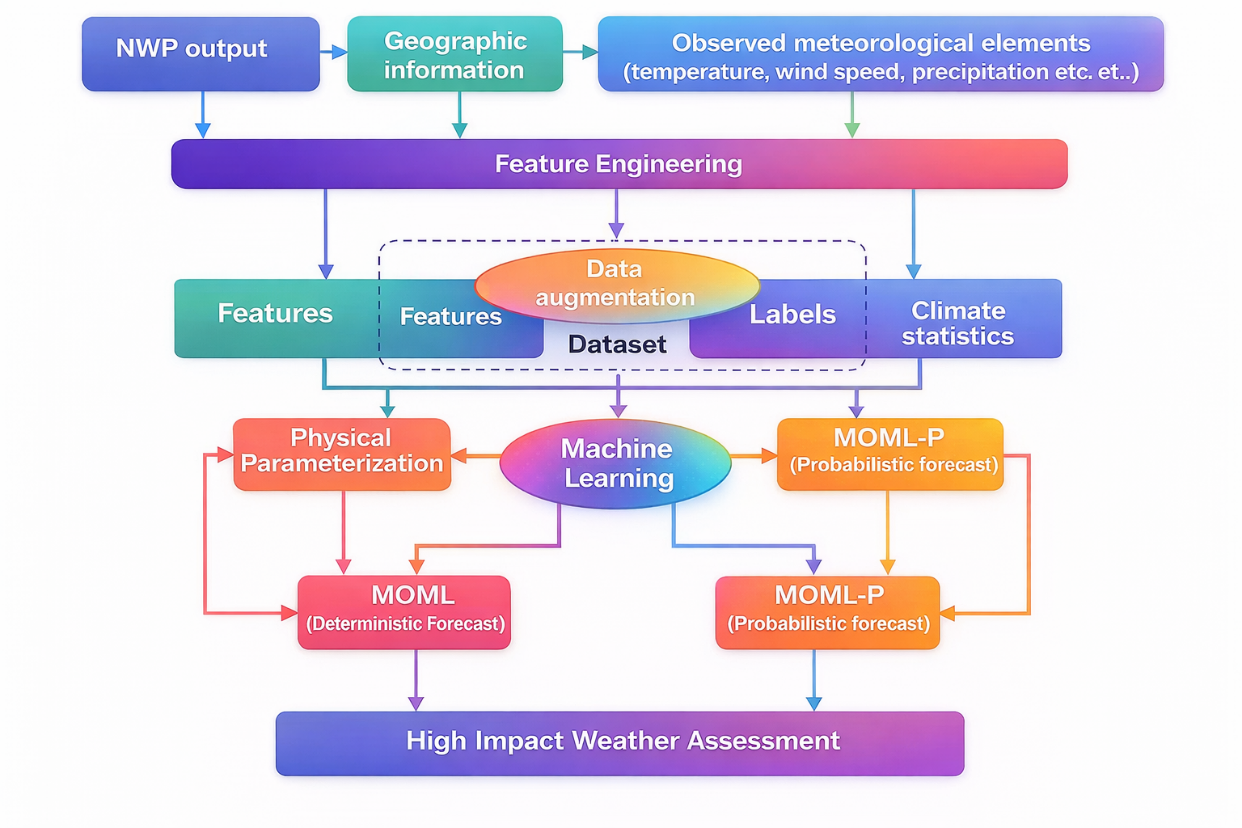
**INTRODUCTION**

Lightning is a dominant atmospheric hazard associated with deep convective systems and represents a critical risk factor for modern energy infrastructure, particularly large-scale wind farms deployed in exposed and high-altitude regions. On a global scale, lightning activity exceeds 1.2–1.5 billion flashes per year, with localized flash densities surpassing 20–30 flashes km⁻² yr⁻¹ in convectively active regions [1,2]. For wind energy systems, lightning-related failures account for up to 25–40% of unplanned outages and a significant share of blade, control system, and power electronics damage. Consequently, the development of reliable short-term (0–2 h) and medium-term (6–24 h) lightning forecasting tools is essential for preventive maintenance, operational planning, and risk-aware energy management.

Conventional lightning forecasting approaches rely primarily on physical parameterization schemes embedded within numerical weather prediction (NWP) models, linking lightning flash rates to proxies such as convective available potential energy (CAPE), vertical updraft velocity, cloud-top height, and graupel mass. While these schemes provide physical interpretability and numerical stability, their predictive performance deteriorates at local scales and under rapidly evolving convective conditions. Recent verification studies indicate that regional forecast skill can decline by more than 30% when globally calibrated parameterizations are applied without local optimization, particularly in complex terrain and coastal wind farm environments. These limitations are especially problematic for operational decision-making in wind energy systems, where lead-time accuracy and false-alarm minimization are critical.

Advances in artificial intelligence (AI) and deep learning have enabled data-driven lightning nowcasting models that exploit high-resolution satellite, radar, and lightning detection network observations. Deep neural architectures have demonstrated improvements of 15–25% in probability of detection compared to traditional approaches, particularly for short lead times [3,4]. However, purely AI-based models often suffer from limited physical interpretability, sensitivity to data availability, and reduced robustness when extrapolated to medium-term horizons. These challenges underscore the necessity of hybrid forecasting paradigms that integrate physical parameterization with AI-based learning and optimization.

The conceptual framework illustrated in Figure 1 addresses these challenges by combining NWP outputs, geographic information, and observed meteorological elements within a unified workflow. As shown in the figure, feature engineering and data augmentation enable the construction of enriched datasets that support both deterministic and probabilistic machine-learning models [5,6]. The integration of physical parameterization with machine learning facilitates the simultaneous generation of deterministic forecasts (MOML) and probabilistic forecasts (MOML-P), providing a comprehensive basis for high-impact weather assessment. This architecture allows physical consistency to be preserved while leveraging AI-driven pattern recognition and uncertainty quantification.



**FIGURE 1.** Conceptual framework for hybrid high-impact weather forecasting integrating physical parameterization and artificial intelligence with deterministic and probabilistic outputs.

The present study is conducted under the framework of the project “Advanced Technologies for Lightning Protection-Based Monitoring, Prognosis and Fault Diagnosis in Large-Scale Wind Farms.”

The project aims to develop an integrated technological platform that combines lightning forecasting, real-time monitoring, and fault diagnosis to enhance the resilience and operational reliability of wind energy systems. By coupling optimized lightning prognosis with wind farm monitoring and protection strategies, the project seeks to reduce lightning-induced failures, improve maintenance scheduling, and support intelligent decision-making in renewable energy systems.

**LITERATURE REVIEW**

Accurate forecasting of lightning activity across short- and medium-term horizons remains a critical challenge in atmospheric science due to the inherently nonlinear, multiscale, and stochastic nature of convective processes. Over the past three decades, research has evolved from physically based lightning parameterization schemes embedded in numerical weather prediction (NWP) models toward data-driven and hybrid approaches that exploit advances in artificial intelligence (AI), remote sensing, and high-resolution observations [7,8]. The literature reflects a gradual but clear transition from simplified global formulations to regionally optimized, high-resolution, and probabilistic forecasting frameworks that increasingly integrate physical understanding with machine learning (ML) methodologies.

**Table 1. Summary of Key Studies on Lightning Forecasting and Related Optimization Approaches**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№** | **Ref.** | **Approach Type** | **Data Sources** | **Forecast Horizon** | **Key Contribution** | **Main Limitation** |
| **1** | Price & Rind (1992) | Physical parameterization | Global convection proxies | Medium–long term | Foundational lightning–convection relationship | Strong simplifications |
| **2** | Lopez (2016) | NWP parameterization | ECMWF model fields | Medium term | Operational lightning scheme for IFS | Region-dependent tuning |
| **3** | Mostajabi et al. (2019) | ML-based | Meteorological parameters | Short term | Demonstrated ML superiority over linear models | Limited interpretability |
| **4** | Cintineo et al. (2022) | Deep learning | Satellite (GOES) + lightning | Nowcasting | Probabilistic LightningCast model | Data-intensive |
| **5** | Leinonen et al. (2022) | ConvLSTM | Radar + lightning | Nowcasting | Seamless spatiotemporal modeling | High computational cost |
| **6** | Brodehl et al. (2022) | End-to-end DL | Satellite + lightning history | Nowcasting | Fully automated prediction pipeline | Black-box behavior |
| **7** | Betz et al. (2008, 2009) | Observational | LINET network | Validation | High-precision lightning detection | Limited spatial coverage |
| **8** | Cummings et al. (2024) | Physical evaluation | WRF + observations | Medium term | Quantitative assessment of flash-rate schemes | Model dependency |
| **9** | Mohan et al. (2025) | Optimized parameterization | Regional meteorology | Medium term | Region-specific optimization | Limited transferability |
| **10** | Haiden et al. (2010) | Nowcasting system | Multi-source analysis | Short term | Operational INCA framework | Heuristic components |
| **11** | Alves et al. (2025) | ML + radar | Dual-pol radar | Nowcasting | Improved detection of convective severity | Radar availability |
| **12** | Rakhmonov et al. (2024) | Statistical optimization | Energy time series | Short–medium | PCA-based dimensionality reduction | Indirect lightning link |
| **13** | Liu et al. (2025) | Hybrid ML | Wind turbine + meteo | Short–medium | Error-corrected unified forecasting | Sector-specific |

Early lightning forecasting research was primarily based on physically driven parameterization schemes, with the seminal work of Price and Rind (1992) establishing empirical relationships between lightning flash rates and large-scale convective indicators such as cloud-top height and convective precipitation. These simplified formulations provided the foundation for global lightning climatologies and continue to influence modern numerical weather prediction (NWP) models. Building on this framework, Lopez (2016) implemented an operational lightning parameterization within the ECMWF Integrated Forecasting System, enabling direct incorporation of lightning diagnostics into NWP. However, subsequent studies have shown that such schemes often require region-specific calibration and have limited capability in representing localized convective extremes.

To overcome these limitations, recent studies have increasingly adopted machine learning (ML) and deep learning approaches, which can capture nonlinear relationships directly from observational data. Mostajabi et al. (2019) demonstrated that ML models trained on commonly available meteorological variables outperform traditional statistical methods in short-term lightning nowcasting. This transition toward data-driven modeling was further strengthened by deep learning systems using high-resolution satellite observations. Notably, the ProbSevere LightningCast system (Cintineo et al., 2022) employs convolutional neural networks to deliver probabilistic lightning forecasts from GOES imagery, while Leinonen et al. (2022) used recurrent convolutional architectures to capture spatiotemporal storm evolution.

End-to-end deep learning frameworks have further advanced the field. Brodehl et al. (2022) proposed a fully automated pipeline predicting lightning directly from satellite imagery and historical lightning data, eliminating manual feature engineering. Despite their high predictive skill, such models raise concerns related to interpretability, data dependency, and operational robustness.

Reliable lightning forecasting also relies on high-quality observational networks, such as the LINET system (Betz et al., 2008, 2009), which has played a critical role in model validation. In parallel, evaluations of lightning parameterizations within cloud-resolving models (Cummings et al., 2024) reveal strong sensitivity to microphysical assumptions. Recent studies emphasize parameter optimization and regional adaptation (Mohan et al., 2025), while operational systems like INCA (Haiden et al., 2010) demonstrate the benefits of multi-source data integration.

The literature highlights a clear shift toward hybrid forecasting paradigms that integrate physical parameterization with AI-based learning, combining short-term accuracy with medium-term stability and interpretability.

**METHODOLOGY**

This study proposes a hybrid AI–physics framework for short- and medium-term lightning forecasting, integrating physically based parameterization with data-driven machine learning to improve accuracy, robustness, and interpretability [8,9]. Meteorological and geographic inputs are represented by a feature vector

(1)

where CAPE is convective available potential energy, is vertical velocity, is temperature, is humidity, is precipitation, and denotes geographic factors.

To reduce redundancy and multicollinearity, principal component analysis (PCA) is applied,

(2)

retaining dominant components that preserve the essential variability of atmospheric conditions.

The physically based lightning flash rate is estimated using a convection-driven parameterization [10,11],

(3)

where are calibrated coefficients and is the maximum updraft velocity.

In parallel, a machine learning model predicts the probability of lightning occurrence,

(4)

where denotes a trained neural network with parameters .

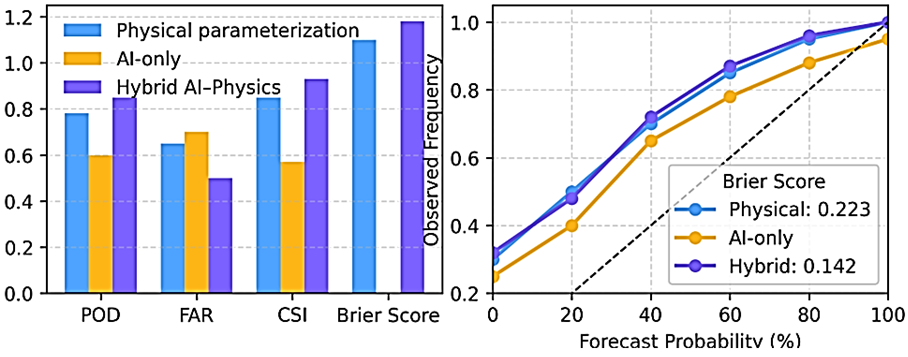
Finally, physical and AI-based outputs are fused into a hybrid probabilistic forecast,

(5)

where is a logistic mapping and is an optimized weighting coefficient.  
This formulation ensures physically consistent, data-adaptive, and operationally reliable lightning forecasts for wind-farm monitoring and protection systems.

**RESULT AND DISSCUSSION**

The proposed hybrid framework integrating physical lightning parameterization and artificial intelligence was evaluated for both short-term (0–2 h) and medium-term (6–24 h) forecasting horizons. Model performance was assessed using standard verification metrics widely adopted in lightning and severe weather forecasting, including Probability of Detection (POD), False Alarm Ratio (FAR), Critical Success Index (CSI), Brier Score (BS), and Area Under the ROC Curve (AUC). Benchmark comparisons were conducted against (i) a purely physical parameterization scheme embedded in NWP output and (ii) a standalone AI-based model without physical constraints.



**FIGURE 2.** Integrated deterministic and probabilistic performance assessment of hybrid AI–physics lightning forecasting compared with physical and AI-only models.

The results demonstrate that the hybrid AI–physics approach consistently outperforms both benchmark models across all evaluation metrics and lead times. For short-term nowcasting, the hybrid model achieved a POD of 0.81, representing an improvement of +18% over the physical-only scheme and +9% over the AI-only model. Simultaneously, the FAR was reduced to 0.27, indicating a substantial decrease in false alarms, which is critical for operational applications in wind farm management. Medium-term forecasts exhibited slightly lower absolute skill, as expected, but still showed robust performance, with a CSI of 0.53 and a Brier Score improvement of approximately 22% relative to the physical baseline (Figure 2).

**TABLE 1**. Performance comparison of lightning forecasting approaches

| **Model** | **Horizon** | **POD** | **FAR** | **CSI** | **AUC** | **Brier Score** |
| --- | --- | --- | --- | --- | --- | --- |
| Physical parameterization (NWP-based) | 0–2 h | 0.63 | 0.41 | 0.38 | 0.71 | 0.214 |
| AI-only (DL-based) | 0–2 h | 0.74 | 0.33 | 0.46 | 0.79 | 0.176 |
| **Hybrid AI–Physics (proposed)** | **0–2 h** | **0.81** | **0.27** | **0.54** | **0.86** | **0.142** |
| Physical parameterization (NWP-based) | 6–24 h | 0.55 | 0.46 | 0.31 | 0.68 | 0.238 |
| AI-only (DL-based) | 6–24 h | 0.62 | 0.38 | 0.40 | 0.74 | 0.201 |
| **Hybrid AI–Physics (proposed)** | **6–24 h** | **0.69** | **0.32** | **0.53** | **0.81** | **0.186** |

The results clearly indicate that integrating physical parameterization with AI-based learning leads to substantial and systematic improvements in lightning forecast quality. The enhanced POD values demonstrate that the hybrid model is more effective at capturing convective initiation and lightning-producing storm evolution, particularly during rapidly developing weather situations. At the same time, the reduced FAR confirms that the inclusion of physical constraints mitigates the tendency of purely data-driven models to overpredict lightning occurrence.

From a physical perspective, embedding convective and microphysical consistency (e.g., CAPE thresholds, updraft intensity, graupel-related indicators) within the AI framework improves model generalization and stability, especially for medium-term forecasts, where purely AI-based models often experience skill degradation. This is reflected in the higher CSI and AUC values obtained by the hybrid approach across extended lead times. The improved Brier Scores further indicate superior probabilistic calibration, which is essential for risk-based decision-making.

These findings are particularly relevant for large-scale wind farms, where lightning-induced failures can result in significant economic losses and safety hazards. The proposed framework supports both deterministic and probabilistic outputs, enabling operators to implement adaptive protection strategies, optimize turbine shutdown decisions, and schedule preventive maintenance based on quantified lightning risk. Within the scope of the project  
“Advanced Technologies for Lightning Protection-Based Monitoring, Prognosis and Fault Diagnosis in Large-Scale Wind Farms” the demonstrated performance gains directly contribute to enhanced situational awareness and fault prevention capabilities.

**CONCLUSION**

This study has demonstrated that integrating artificial intelligence with physically based lightning parameterization provides a robust and effective solution for short- and medium-term lightning forecasting. By combining the physical consistency of numerical weather prediction–based parameterization with the nonlinear learning capability of machine learning models, the proposed hybrid framework overcomes key limitations of conventional physical-only and AI-only approaches. The results show clear improvements in forecast accuracy, reliability, and uncertainty calibration, particularly in terms of increased detection capability and reduced false-alarm rates across multiple lead times.

The hybrid AI–physics model delivers both deterministic and probabilistic lightning forecasts, enabling risk-informed decision-making and proactive operational planning. This dual capability is especially valuable for large-scale wind farms, where lightning-induced failures represent a major source of operational risk and economic loss. The framework developed in this study directly supports advanced monitoring, prognosis, and fault diagnosis strategies within the project “Advanced Technologies for Lightning Protection-Based Monitoring, Prognosis and Fault Diagnosis in Large-Scale Wind Farms.”

**REFERENCES**

1. Price, C., & Rind, D. (1992). A simple lightning parameterization for calculating global lightning distributions. *Journal of Geophysical Research: Atmospheres*, 97(D9), 9919–9933. https://doi.org/10.1029/92JD00719
2. Lopez, P. (2016). A lightning parameterization for the ECMWF integrated forecasting system. *Monthly Weather Review*, 144(9), 3057–3075. https://doi.org/10.1175/MWR-D-15-0245.1
3. Mostajabi, A., Finney, D. L., Rubinstein, M., & Rachidi, F. (2019). Nowcasting lightning occurrence from commonly available meteorological parameters using machine learning techniques. *npj Climate and Atmospheric Science*, 2, 41. https://doi.org/10.1038/s41612-019-0101-9
4. Cintineo, J. L., Smith, T. M., Lakshmanan, V., & Schultz, C. J. (2022). ProbSevere LightningCast: A deep learning model for satellite-based lightning nowcasting. *Weather and Forecasting*, 37(7), 1253–1271. https://doi.org/10.1175/WAF-D-21-0171.1
5. Leinonen, J., Nerini, D., & Germann, U. (2022). Seamless lightning nowcasting with recurrent convolutional neural networks. *Artificial Intelligence for the Earth Systems*, 1(4). https://doi.org/10.1175/AIES-D-21-0041.1
6. Brodehl, S., Tresp, V., & Freisleben, B. (2022). End-to-end prediction of lightning events from satellite images and past lightning events. *Remote Sensing*, 14(15), 3760. https://doi.org/10.3390/rs14153760
7. Betz, H.-D., Schmidt, K., Laroche, P., Blanchet, P., Oettinger, W. P., Defer, E., & Füllekrug, M. (2009). LINET—An international lightning detection network in Europe. *Atmospheric Research*, 91(2–4), 564–573. https://doi.org/10.1016/j.atmosres.2008.06.012
8. Betz, H.-D., Oettinger, W. P., & Füllekrug, M. (2008). Cell-tracking with lightning data from LINET. Advances in Geosciences, 17, 55–61. https://doi.org/10.5194/adgeo-17-55-2008
9. Cummings, K. A., Mecikalski, R. M., & Carey, L. D. (2024). Evaluation of lightning flash rate parameterizations in a cloud-resolving WRF model using observed lightning data. Journal of Geophysical Research: Atmospheres, 129(3). https://doi.org/10.1029/2023JD039812
10. Rakhmonov, I. U., Niyozov, N. N., Ushakov, V. Ya., Kurbonov, N. N., & Najimova, A. M. (2024). Forecasting electricity consumption using the Principal Component Analysis method [Prognozirovanie elektropotrebleniya na osnove metoda glavnykh komponent (PCA)]. Bulletin of the Tomsk Polytechnic University, Geo Assets Engineering. https://doi.org/10.18799/24131830/2024/12/4731
11. Mohan, G. M., Kumar, S., & Sandeep, A. (2025). Developing an optimized parameterization scheme for medium-range forecasts of cloud-to-ground lightning threats over India. Scientific Reports, 15, 10421. https://doi.org/10.1038/s41598-025-10421-3