Research Roadmap for Next-Generation Quantitative Precipitation Nowcasting: Transitioning from Modular to Interpretable Impact-Based Deep Learning

I. Foundational Analysis of the Existing Nowcasting Framework

Highly short-term forecasting, known as nowcasting (forecasts up to a few hours ahead), is a critical endeavor, particularly for mitigating the severe adverse impacts associated with rapidly evolving mesoscale convective systems (MCSs). Hazardous meteorological conditions, such as flash flooding, frequently occur within hours of these rainstorm events, necessitating timely and high-resolution warnings to reduce potential flood exposure in communities.

1.1. Context and Significance of Heavy Rainfall Nowcasting (QPN)

1.1.1. Limitations of Traditional NWP in High-Resolution Nowcasting

Traditional numerical weather prediction (NWP) models, typically employed for daily weather forecasting, face inherent limitations when adopted for high spatial (e.g., 1 km) and temporal (e.g., 5 min) resolution applications. These limitations stem from deficiencies in determining precise initial and boundary conditions, requiring long spin-up times, and confronting challenges in physical parameterization. Consequently, traditional NWP models struggle to provide useful nowcasting products for high-resolution applications such as hydrometeorological modeling in urban or catchment areas.

This functional gap validates the use of data-driven approaches. Recent studies have demonstrated the compelling necessity of deep learning (DL) solutions, showing that state-of-the-art physics-embedded deep generative models, such as NowcastNet, significantly outperform the latest generation of NWP models, including the High-Resolution Rapid Refresh (HRRR) model, specifically for extreme precipitation events exceeding 16 \text{ mm/h}. This demonstrated superiority justifies the fundamental research focus on refining DL-based solutions for hazard forecasting.

1.1.2. The Role of Machine Learning in Operational QPN

Machine learning methods are strategically applied to overcome NWP limitations by leveraging the high-resolution data available from remote sensors, such as weather radars and satellites. By training ML models on empirical data, it is possible to interpolate and extrapolate rainfall patterns rapidly given new measured input variables. The optimal selection of architecture is crucial for success: ML models based on convolutional computations are particularly adept at capturing static spatial correlations, while recurrent network-based methods are specifically capable of modeling temporal trends within the data sequence. A robust body of research confirms the potential utility of various ML architectures, including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and Deep Generative Modeling (DGM), in accurately nowcasting rainfall.

1.2. Deconstruction of the Target Model (Wang et al., 2023)

The foundational study utilized an ensemble of machine learning regression and classification techniques to quantify the relative importance of radar variables for nowcasting heavy rainfall over the Korean Peninsula.

1.2.1. Dual-Polarimetric Radar Features and Preparation

The modeling relies on data products derived from a dual-polarimetric Doppler radar, focusing on variables aggregated at the storm cell level. Over 15 variables were investigated, including geometric properties like the longest radius (R_{mj}) and shortest radius (R_{mn}), kinematic information (U, V, Direction), and critical intensity metrics such as Storm Area Averaged Reflectivity (MeanZ), Volume, and Maximum Vertically Integrated Liquid Water (Max VIL). Data preparation involved standardizing the variables, converting them to an integer format, and applying base-10 logarithmic transformations to highly positively skewed variables to enhance their representativeness and stabilize their variation prior to model training.

1.2.2. Multistage Architecture

The framework operates in a distinct modular fashion, which is a key differentiator from modern end-to-end deep learning approaches.
The methodology comprises three stages. The first stage involves **Tracking & Classification**, utilizing the fuzzy logic algorithm for storm tracking (FAST) to identify and categorize storm cells into types like Convective Cell (CC) and various squall lines. These were simplified into two major categories for analysis: CC and Mesoscale Squall Line (MSL) datasets. The second stage is **Regression**, where traditional ML methods (Lasso, RFR, SVR, and ANN) predicted cell-averaged Mean Rain Rate and Top 10% Mean Rain Rate for lead times of 30 and 60 minutes. A permutation importance algorithm was integral to this stage for variable selection. The final stage is **Warning Issuance**, where the recalibrated ML regression models are coupled with separate CNN classification models to predict storm cell locations and issue an alert if the predicted Top 10% Mean Rain Rate meets or exceeds 30 mm/h.

1.3. Critical Review of Findings and Identified Research Gaps

1.3.1. Quantification of Feature Importance: Key Variables and Consistency

The study successfully quantified feature importance, revealing that radar variables associated with water mass and intensity were highly important predictors of heavy rainfall. These consistently pertinent predictors included storm area volume, storm area maximum vertically integrated liquid water (Max VIL), storm area averaged reflectivity (MeanZ), echo bottom height

(Base), and longest radius (R_{mj}).
Conversely, kinematic variables—the zonal (U) and meridional (V) components of cell motion velocity, as well as storm propagation direction—were ranked as the least important predictors of rainfall rate. This observation indicates that for very short lead times (30–60 minutes), the instantaneous internal physical state of the storm (intensity and volume) is statistically more influential in predicting the future rainfall *rate* than its direction of movement. However, this raises an operational concern: while storm velocity may be statistically trivial for rate prediction, slow-moving heavy storms are critical factors in flood management. This highlights that the model, as designed, lacks the inherent operational physics necessary to prioritize the hazard implications of storm kinematics.

1.3.2. Performance Disparities by MCS Storm Type

A critical limitation identified was the performance disparity based on MCS type. The classification models for Convective Cells (CC), which are generally smaller and more difficult to track accurately, exhibited the poorest performance, achieving high recall but notably low precision, resulting in the lowest F1 scores.
The combination of high recall (correctly identifying most positive instances) and low precision (a high rate of false positives) for CC models implies that the storm cell boundary defined by the FAST algorithm and the use of cell-averaged features are too coarse for accurately localizing small, rapidly evolving cells. This results in predicted storm areas that are too large relative to the actual observed storm cell areas. The base paper concludes that future work must incorporate MCS type information to improve ML modeling, which directs the necessary research toward highly granular, pixel-wise conditional modeling to enhance prediction precision and reduce the frequency of false alarms.

1.3.3. Limitations in Warning System Precision and Spatial Blurring

The current warning system issues alerts associated with the "entire predicted area of the storm cell" based on averaged metrics. This system is critically limited by classification precision metrics consistently below 0.7 , resulting in a high rate of false alarms for potential warning locations. Furthermore, the explicit use of the FAST tracking algorithm introduces potential identification and tracking errors when dealing with irregularly shaped storms or when clustered storm segments are mistakenly detected as a single larger cell.
To evolve into a high-precision system, the current brittle, two-step modular structure must be superseded by a unified, end-to-end deep learning architecture. Such a framework must implicitly learn motion and organization , bypassing the tracking errors and overcoming the limitations that result in a blurring effect in the spatial mapping of rain rates, particularly detrimental for longer lead times.

II. Technical Roadmap for Model Upgrade: Integrating State-of-the-Art Deep Learning

The model upgrade must transition from a modular, deterministic approach to a unified, probabilistic, spatio-temporal framework, specifically targeting robust prediction of extreme events and small-scale convection.

2.1. Adopting Modern Spatio-Temporal Deep Learning Architectures

2.1.1. Transition to Unified End-to-End Forecasting

The upgrade mandates the abandonment of the two-step modeling paradigm involving explicit storm tracking (FAST) and discrete regression/classification stages. End-to-end models, such as those based on Convolutional LSTMs (ConvLSTM) or Deep Generative Models (DGM), are preferred because they avoid tracking errors and simultaneously capture spatiotemporal dependencies. The ConvLSTM architecture, which combines CNNs for spatial correlation and recurrent layers for temporal trends, is considered an ideal design for capturing the spatiotemporal associations necessary for nowcasting rainfall.
2.1.2. Deep Generative Models (DGM) and ConvLSTM/U-Net Benchmarking
The **Proposed Architecture** should utilize a hybrid model, such as one integrating the U-Net architecture with ConvLSTM cells. The U-Net's flexible structure excels at multi-resolution feature extraction, minimizing blurring and allowing for easier adjustment to accommodate multiple input streams compared to fixed ConvLSTM structures.
To secure novelty and publication, this model must be rigorously benchmarked against the principles of state-of-the-art DGM techniques (e.g., DGMR or MetNet). These models have demonstrated exceptional skill, often leveraging global NWP and satellite data to produce highly skillful precipitation forecasts over larger regions and extended horizons, reaching up to 12 hours.

2.2. Enhancing Data Integration and Lead Time

To achieve robust performance, especially beyond the initial 60-minute window, the input data must be expanded to provide comprehensive physical and dynamical context.

2.2.1. Expanding Input Features to Leverage Dual-Pol Exclusivity

While the base model used dual-polarimetric radar data, the variables selected did not fully exploit the microphysical insights unique to polarimetry. True advancement requires the integration of variables like **Specific Differential Phase (\text{K}_{\text{DP}})** and **Differential Reflectivity (\text{Z}_{\text{DR}})**. These features provide critical microphysical information concerning hydrometeor shape, size, and phase, which is essential for correcting attenuation and accurately estimating heavy rain rate. These variables specifically offer deep learning models an opportunity to better forecast the organization and dynamic evolution of convective storms.
2.2.2. Fusion of Auxiliary Information for Extended Horizons
Extending the reliable lead time beyond radar extrapolation capabilities requires multi-source data fusion:

- **Lightning Data Integration:** Ground-based lightning detection provides insight into the charge separation processes, serving as an early indicator of severe convection. Integrating lightning, radar, and satellite data enhances the model's ability to capture severe storm microphysics and charge separation.
- **NWP and Satellite Data:** To overcome the inherent limitations of radar-only models regarding short forecast lead times, the model must efficiently integrate physics-based NWP output and geostationary satellite data. This fusion allows the model to capture

long-range interactions, enabling robust probabilistic forecasting over extended lead times up to 8 hours.

2.3. Optimization Strategies for Extreme Rainfall Prediction

The challenge of heavy rainfall underestimation, noted as a difficulty for many DL models , must be addressed directly to ensure accurate warnings based on the 30 \text{ mm/h} threshold.

2.3.1. Mitigating Heavy Rainfall Underestimation Bias

In nowcasting, extreme precipitation events are infrequent minority classes. Standard loss functions, such as Mean Squared Error (MSE), prioritize minimizing average prediction error, inherently biasing the model to perform well on common, low-intensity rainfall while consistently underestimating rare, high-impact extremes.

2.3.2. Implementation of Weighted and Ordinal Loss Functions

To force the model to prioritize accurate prediction of the hazard threshold, specialized loss functions are required:

- **Weighted Loss (e.g., Focal Loss):** Applying a weighted loss function, which significantly increases the penalty for errors made in high-intensity bins, demonstrably improves the accuracy of extreme rainfall prediction. This approach requires meticulous parameter sensitivity analysis to identify optimal weighting strategies tailored to the frequency of heavy rainfall in the study area.
- **Ordinal Consistent Loss:** Since precipitation intensity is fundamentally ordinal, using a loss function that explicitly accounts for the hierarchical rank relationship between precipitation bins (e.g., 10 \text{ mm/h} < 30 \text{ mm/h}) provides a more physically consistent optimization target than using standard classification losses.

2.3.3. Training Schemes for Small-Scale Convective Cell Enhancement

The consistently low precision observed for CC storm cells points to a failure in localizing small-scale convective structures. This can be mitigated by incorporating techniques used in deep generative modeling, such as integrating an adversarial loss component (GAN). This method serves to sharpen small-scale structures in the generated output and reduce the characteristic blurring effect of DL models. Such enhancement is necessary to improve the robustness and practical utility of nowcasting models for small-scale, extreme rainfall phenomena.

2.4. Achieving Probabilistic and Uncertainty-Aware Forecasting

Operational agencies require quantifiable confidence in forecasts. The deterministic output of the base model must be replaced by a comprehensive probabilistic framework.

2.4.1. Shifting to Probabilistic Output Maps

The updated model must generate an estimate of the **probability distribution of precipitation intensities** rather than a single deterministic value. The output should consist of high-resolution probabilistic maps indicating \text{P}(\text{R} > \text{T}) for relevant heavy rainfall thresholds (e.g., \text{T} = 30 \text{ mm/h}).

2.4.2. Quantifying Uncertainty

The generation of consistent probabilistic maps automatically provides robust uncertainty quantification, offering essential high-confidence metrics for real-time warning issuance. Furthermore, the model must formally acknowledge that confidence decreases with lead time. This temporal uncertainty should be addressed during training by implementing **lead time weights** (W_t), normalized from an exponential distribution, which controls the decay rate and determines the relative weighting of the first forecast timesteps over later ones.

III. Advancing Novelty: Interpretation and Impact-Based Forecasting (The "Not Existed Till Date" Roadmap)

The path to high-impact publication requires addressing the dual mandate of scientific transparency and operational effectiveness. Integrating advanced Explainable AI (XAI) and creating a functional Impact-Based Forecasting (IBF) chain represents genuinely novel directions for this specific application.

3.1. Interpretable Deep Learning for Meteorological Validation

The complexity of deep learning models necessitates rigorous explanation to ensure they are trusted and adopted by operational meteorologists.

3.1.1. Implementation of SHAP and LIME for Local Feature Attribution

Advanced post-hoc explanation methods must be applied. Utilizing SHAP (SHapley Additive exPlanations) and SAGE (Shapley Additive Global Explanation) is recommended to assess feature importance both locally and globally. SHAP is mathematically sound for attributing the prediction value to specific input features based on coalitional game theory.
A significant novel application involves using SAGE to quantify the model impact of **feature groups** (e.g., microphysical variables versus kinematic variables). This strategy mitigates the known adverse effects that highly correlated features have on individual feature importance rankings, leading to a more comprehensive and holistic understanding of the model's reliance on physically grouped inputs.

3.1.2. Spatio-Temporal XAI Visualization

Standard scalar feature importance is insufficient for spatio-temporal predictions. Visualization is necessary to understand the model's spatial and temporal attention:

- **Gradient-weighted Class Activation Mapping (Grad-CAM):** Applying Grad-CAM to the ConvLSTM layers reveals the specific spatial regions within the input radar sequence that the network attends to when making a high-intensity prediction. This provides critical visual validation that the model is focusing on physically relevant storm cores, rather than spurious features.

- **Temporal Occlusion:** Systematically removing or corrupting input data at specific prior time steps (temporal occlusion) quantifies the decay of temporal influence. This directly determines how much the model relies on the current state versus observations from earlier time steps, thereby separating the influence of persistence/advection from dynamical evolution learning.

3.1.3. Leveraging XAI to Validate MCS-Type Dependence

The base model observed distinct variable importance rankings based on MCS type. XAI provides the essential tool to scientifically justify this difference. For example, Grad-CAM visualizations must confirm that the attention mechanism for Squall Lines (MSL) heavily weights the long-axis features and spatial organization geometry, while the CC model focuses intensely on the localized core intensity features (Max VIL). This demonstrated evidence of physically consistent, mode-specific feature utilization, validated through XAI, constitutes a major scientific contribution.

3.2. High-Resolution Impact-Based Warning Systems (IBF)

The most transformative and publishable step is the creation of a functional Impact-Based Forecasting (IBF) system that moves beyond the coarse meteorological warning to a quantified societal hazard forecast.

3.2.1. Coupling QPN with Hydrological and Hydrodynamic Models (QPN-H-HD Chain)

The probabilistic QPN ensemble must seamlessly drive an integrated forecasting chain:
The QPN output serves as the input forcing for a high-resolution Hydrological (H) model (e.g., mHM) to calculate surface runoff, which in turn drives the Hydrodynamic (HD) model (e.g., RIM2D) to simulate water movement and depth. This coupled chain offers a physical pathway to quantify impact, unlike the purely empirical approach of the base model.

3.2.2. Hyper-Resolution Mapping of Inundation

Leveraging high-resolution Digital Elevation Models (DEMs), the IBF system must generate output far superior to the current cell-averaged warning. The system must produce hyper-resolution information (e.g., 10-meter grid) on **water depth**, **inundation area**, and critically, identify specific **building footprints likely to be inundated**. This provides actionable, hyper-localized warnings essential for flood preparedness and rapid response.

3.2.3. Developing an Optimal Impact Index

The warning must be based on a quantified impact index rather than a simple rain rate threshold. This index should integrate: (1) The probability of the rainfall event (from the QPN output), (2) the magnitude of the resulting physical hazard (water depth and inundation area calculated by the H-HD chain), and (3) exposure data (locations of critical infrastructure, high-risk areas). This framework ensures that warnings quantify the size and nature of actual impacts, optimizing response efforts.

3.3. Operationalization and Publication Strategy

3.3.1. Benchmarking Protocol

Rigorously proving the updated model's performance is essential. Validation must focus on skill scores relevant to extremes:

- **Metrics for Extremes:** The primary performance metric must be the **Critical Success Index (CSI)** for high precipitation thresholds (e.g., 30 \text{ mm/h}), which measures the accuracy of predicted heavy rainfall events.
- **Comparison:** Performance must be benchmarked against standard persistence methods, extrapolation-based methods (if relevant regionally), and the latest high-resolution regional NWP models. Studies have shown that rigorous comparison against best-in-class NWP (like HRRR) using CSI provides compelling evidence of scientific advancement.

3.3.2. Publication Narrative

The publication narrative must emphasize the breakthrough components: the superior CSI achieved using weighted loss functions, the scientific transparency provided by Spatio-Temporal XAI validation of physical mechanisms, and the transformative operational impact achieved through the QPN-H-HD coupling to deliver hyper-resolution IBF. This strategy allows the research to emphasize its role in reducing global disparities in forecast quality through the integration of high-resolution remote sensing and advanced machine learning.
Table 3 summarizes the proposed novel objectives and methodologies required to ensure the published work addresses research directions "not existed till date."
Table 3: Roadmap for High-Impact Publication: Focusing on Interpretation and Impact-Based Forecasts

Research Phase	**Novel Objective**	**Core Methodology**	**Key Deliverable for Publication**
Phase A: Spatio-Temporal XAI	Rigorously explain complex DL predictions and validate physical consistency.	SHAP/SAGE for feature attribution; Grad-CAM for spatial attention mapping; Temporal Occlusion.	XAI-validated physical mechanism interpretation; quantification of feature group impact.
Phase B: Conditional Modeling	Optimize performance by dynamically conditioning the model based on predicted MCS mode (CC, SLP, MCC).	Conditional DL architecture or multi-task learning; leverage XAI to justify mode-specific feature weights.	Demonstrable superior performance (\text{CSI}/\text{F}1) for difficult-to-predict CC storms.
Phase C: Hyper-Resolution Impact-Based Forecasting (IBF)	Transition from meteorological warning to real-world hazard prediction.	Couple QPN probabilistic ensemble with high-resolution Hydrological (H) and Hydrodynamic (HD)	Near-real-time, hyper-resolution (10m) forecasts of building inundation probability and water depth.

Research Phase	Novel Objective	Core Methodology	Key Deliverable for Publication
		models.	
Phase D: Operational Benchmarking	Prove superior operational value against the established industry benchmarks.	Comparison of \text{CSI} for 30 \text{ mm/h} events against regional NWP models.	Evidence of reducing the global disparity in forecast quality for data-sparse regions.

Works cited

1. Hybrid physics-AI outperforms numerical weather prediction for extreme precipitation nowcasting - PMC - PubMed Central, https://pmc.ncbi.nlm.nih.gov/articles/PMC11573709/ 2. [2510.13050] An Operational Deep Learning System for Satellite-Based High-Resolution Global Nowcasting - arXiv, https://arxiv.org/abs/2510.13050 3. NowDeepN: An Ensemble of Deep Learning Models for Weather Nowcasting Based on Radar Products' Values Prediction - MDPI, https://www.mdpi.com/2076-3417/11/1/125 4. Accurate and Clear Quantitative Precipitation Nowcasting Based on a Deep Learning Model with Consecutive Attention and Rain-Map Discrimination in - AMS Journals, https://journals.ametsoc.org/view/journals/aies/1/3/AIES-D-21-0005.1.xml 5. Nowcasting Heavy Rainfall With Convolutional Long Short-Term Memory Networks: A Pixelwise Modeling Approach - IEEE Xplore, https://ieeexplore.ieee.org/iel7/4609443/10330207/10487101.pdf 6. Improving Nowcasting of Convective Development by Incorporating Polarimetric Radar Variables Into a Deep‐Learning Model - the NOAA Institutional Repository, https://repository.library.noaa.gov/view/noaa/54300/noaa_54300_DS1.pdf 7. RainPro-8: An Efficient Deep Learning Model to Estimate Rainfall Probabilities Over 8 Hours, https://arxiv.org/html/2505.10271v2 8. Lightning Nowcasting Using Dual-Polarization Weather Radar and Machine Learning Approaches: Evaluation of Feature Engineering Strategies and Operational Integration - MDPI, https://www.mdpi.com/2225-1154/13/8/168 9. Seamless Lightning Nowcasting with Recurrent-Convolutional Deep Learning - AMS Journals, https://journals.ametsoc.org/view/journals/aies/1/4/AIES-D-22-0043.1.pdf 10. Enhancing Extreme Rainfall Nowcasting with Weighted Loss Functions in Deep Learning Models - CO Meeting Organizer, https://meetingorganizer.copernicus.org/EGU25/EGU25-19416.html?pdf 11. Unveiling the Role of Weighted Loss Functions in Deep Learning-Based Nowcasting of Extreme Rainfall Events - IEEE Xplore, https://ieeexplore.ieee.org/iel8/36/10807682/11096925.pdf 12. [Literature Review] RainPro-8: An Efficient Deep Learning Model to Estimate Rainfall Probabilities Over 8 Hours, https://www.themoonlight.io/en/review/rainpro-8-an-efficient-deep-learning-model-to-estimate-rainfall-probabilities-over-8-hours 13. Generative Ensemble Deep Learning Severe Weather Prediction from a Deterministic Convection-Allowing Model in - AMS Journals, https://journals.ametsoc.org/view/journals/aies/3/2/AIES-D-23-0094.1.xml 14. (PDF) Focal Frame Loss: A Simple but Effective Loss for Precipitation Nowcasting, https://www.researchgate.net/publication/362336456_Focal_Frame_Loss_A_Simple_but_Effective_Loss_for_Precipitation_Nowcasting 15. STAA: Spatio-Temporal Attention Attribution for Real-Time Interpreting Transformer-based Video Models - arXiv, https://arxiv.org/html/2411.00630v1 16. A Machine Learning Explainability Tutorial for Atmospheric Sciences - AMS Journals, https://journals.ametsoc.org/view/journals/aies/3/1/AIES-D-23-0018.1.pdf 17. LIME vs SHAP: A Comparative Analysis of Interpretability Tools - MarkovML,

https://www.markovml.com/blog/lime-vs-shap 18. Interpretation of a Machine Learning Model for Short-Term High Streamflow Prediction, https://www.mdpi.com/2673-4834/6/3/64 19. Deep Learning for Short-Term Precipitation Prediction in Four Major Indian Cities: A ConvLSTM Approach with Explainable AI - arXiv, https://arxiv.org/html/2511.11152v1 20. Implementing Multi-Hazard Impact-based Forecast and Warning Services - Global Facility for Disaster Reduction and Recovery (GFDRR), https://www.gfdrr.org/sites/default/files/publication/Report%20on%20Implementing%20Multi%20Hazard%20Impact%20Based%20Forecast%20and%20Warning%20Services%2017-02-10%20Final.pdf 21. The future of high-resolution impact-based flood early warning systems, https://agu.confex.com/agu/agu24/meetingapp.cgi/Paper/1592448