PINN-Infused Hybrid ML Forecasting on Lake-Effect Precipitation

Wuyang Zhang[1*], Zhen Luo[2], Jianming Ma[2], Wangming Yuan[3], Tongyu Zhang[4] and Chengwei Feng[5]

[1]Department of Electrical and Computer Engineering, University of Massachusetts Amherst, Amherst, Massachusetts, United States

[2]Department of Computer Sys. Engineering, Northeastern University, Boston, Massachusetts, United States

[3]Department of Computer Science, George Mason University, Fairfax, Virginia, United States

[4]Department of Pathobiology, University of Illinois Urbana-Champaign, Urbana, Illinois, United States

[5]Department of Computer & Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand

*Corresponding author: doggo@ieee.org

Abstract

Lake-effect snow poses severe risks to communities around the Great Lakes. However, accurate prediction remains elusive due to a fundamental challenge: critical satellite observations are unavailable at night when these systems rapidly intensify. We propose a novel approach to lake-effect snow forecasting. First, we solve the temporal data discontinuity problem. Then, we leverage complete observations for physics-informed prediction. Our two-stage framework uses PatchGAN to synthesize missing visible and near-infrared satellite imagery from continuous infrared data. This approach improves forecast accuracy by 59% compared to models trained on incomplete observations. These synthesized sequences then feed into a physics-informed neural network architecture that modifies MetNet-3 and enforces atmospheric conservation laws while processing high-density weather station data at adaptive resolutions. Most remarkably, our approach reveals that harsh lake-effect events become more predictable over longer time periods, improving from 27.1% accuracy at 24 hours to 77.6% at 72 hours as large-scale precursor patterns emerge in the complete observational record. When evaluated using 11 years of Great Lakes data, our framework achieves an overall accuracy of 87.4% for 24-hour forecasts and 81.3% for 72-hour forecasts. This substantially outperforms traditional NWP models (42.3%, 66.5%) and standard deep learning approaches (45.3%, 64.1%). By showing that intelligent data synthesis can unlock the potential of physics-informed machine learning, our work establishes new groundwork for predicting localized severe weather phenomena, which have historically been limited by observational gaps.

Index Terms— Physics-Informed Neural Networks, Lake-Effect Snow Prediction, Cross-Spectral Image Synthesis, Temporal Data Completion, Multi-Scale Meteorological Forecasting, Generative Adversarial Networks, Adaptive Resolution Targeting, ConvLSTM

1 Introduction

Lake-effect snow exemplifies the challenge of predicting localized severe weather in an era of climate extremes. These phenomena occur when Arctic air masses traverse the relatively warm waters of the Great Lakes, undergoing rapid transformation that produces intense, narrow bands of snowfall capable of depositing over 100 cm in 48 hours (Figure 1). The December 2022 Buffalo snowstorm, which resulted in 47 deaths, underscores the critical need for an accurate prediction of these events [26]. However, despite decades of research and advances in weather modeling, lake-effect snow remains notoriously difficult to forecast because of a fundamental observational challenge: the very data needed to track these rapidly evolving systems become unavailable precisely when the systems are most active.

The core challenge lies in the temporal discontinuity of satellite observations. Visible and near-infrared imagery provides crucial information about cloud structure and evolution, yet these spectral bands are only available during daylight hours, approximately 7-8 hours during winter months when lake-effect snow is most prevalent. This creates critical 12- to 16-hour gaps in observations, often coinciding with evening and early morning periods when cold air advection intensifies and lake effect systems rapidly develop [18]. Current forecasting approaches attempt to work around these gaps through various strategies. Numerical Weather Prediction (NWP) models rely on sparse ground observations and coarse-resolution physics simulations, while machine learning methods simply skip over missing timesteps. Neither approach adequately captures the continuous evolution of atmospheric processes that drive lake-effect formation.

This observational discontinuity cascades into two additional challenges that have limited prediction accuracy. First, without continuous monitoring, the models cannot capture the mesoscale processes (atmospheric phenomena occurring at scales of 2-200 km) that organize scattered convection into coherent snow bands. These bands, typically 10-20 km wide, fall below the resolution of operational NWP models (10-25 km) and require persistent tracking to predict their formation, movement, and intensification [22]. Second, the lack of complete temporal data prevents the models from learning the physical relationships between precursor atmospheric conditions and subsequent precipitation. Although physics-based

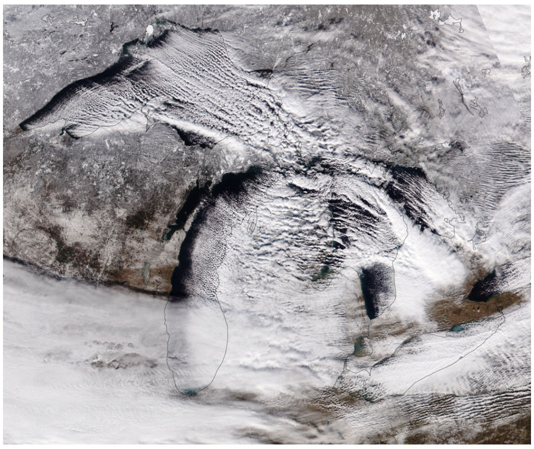

Figure 1: Satellite imagery capturing intense lake-effect snow bands flowing off the Great Lakes. These narrow bands, typically 10-20 km wide, can produce dramatically different conditions in neighboring communities—heavy snowfall in one location while areas just kilometers away remain clear.

models encode these relationships through equations, they struggle with nonlinear lake-atmosphere interactions; conversely, data-driven models could potentially learn these complex patterns but require continuous observations to do so effectively [1, 21].

Our Approach: Data Synthesis Enables Physics-Informed Prediction These fundamental limitations motivate a paradigm shift in how we approach lake-effect snow forecasting. Rather than developing increasingly sophisticated models to work around observational gaps—the traditional approach that has yielded incremental improvements over decades—we propose addressing the root cause directly. We hypothesize that solving the data completeness problem first will unlock the full potential of physics-informed machine learning approaches that have been constrained by fragmented observations.

We propose a new approach to lake-effect snow prediction: rather than working around observational gaps, we first solve the data completeness problem through intelligent synthesis, then leverage these complete data for physics-informed prediction. Our approach introduces a two-stage framework that fundamentally reimagines how we handle missing meteorological observations. In the first stage, we employ PatchGAN (a type of Generative Adversarial Network that operates on image patches rather than whole images), to synthesize missing visible and near-infrared imagery from the continuously available infrared band. Unlike simple interpolation, our approach learns the complex physical relationships between spectral signatures, cloud properties, and atmospheric states, generating meteorologically consistent imagery that maintains the spatial and temporal coherence necessary for tracking lake-effect development. This synthesis transforms fragmented observations into continuous 15-minute interval sequences that span complete diurnal cycles.

The second stage leverages these temporally complete observations within a novel prediction architecture that combines the pattern recognition capabilities of deep learning with the physical constraints of atmospheric science. We enhance the MetNet-3 architecture (a state-of-the-art neural weather model from Google DeepMind) by replacing its dependency on coarse NWP data with a Physics-Informed Neural Network (PINN) module—a neural network that incorporates physical laws as constraints during training—that processes high-density weather station observations. The framework also employs Convolutional Long Short-Term Memory (ConvLSTM) networks, which are specialized recurrent neural networks that handle spatiotemporal data by replacing standard LSTM's fully connected operations with convolutions to preserve spatial structure while modeling temporal dependencies. This modification enables fine-scale resolution where needed while enforcing fundamental conservation laws, mass continuity, energy balance, and thermodynamic constraints, which ensure that predictions remain physically plausible throughout the 72-hour forecast horizon. To maintain computational efficiency despite the increased resolution, we implement adaptive spatial targeting that dynamically allocates resources based on lake-effect probability, achieving 500-meter resolution in high-risk zones while using coarser grids elsewhere.

The synergy between complete temporal observations and physics-informed prediction yields remarkable improvements in forecast accuracy. Our PatchGAN synthesis achieves a 59% improvement in Critical Success Index (0.67 vs. 0.42) compared to models trained on gapped data, demonstrating that continuous observations are essential for capturing atmospheric evolution. Most surprisingly, our framework shows dramatic improvement in predicting harsh lake-effect events at extended forecast horizons—accuracy increases from 27.1% at 24 hours to 77.6% at 72 hours. This counterintuitive result reveals that severe events are preceded by large-scale atmospheric patterns that become increasingly predictable over multi-day timescales, but only when models have access to complete observational sequences that capture these evolving patterns. Overall, our approach achieves 87.4% accuracy for 24-hour forecasts and maintains 81.3% accuracy at 72 hours, substantially outperforming both physics-based FLake NWP and data-driven MetNet-3 baselines.

Beyond improving lake-effect snow prediction, this work demonstrates the power of addressing fundamental data limitations in environmental forecasting. By solving the temporal completeness problem first, we enable physics-informed deep learning approaches to reach their full potential. The framework's success suggests that many challenging prediction problems in meteorology and related fields may benefit more from intelligent data synthesis than from increasingly complex models trained on incomplete observations. Our approach is particularly relevant as climate change intensifies extreme weather events, demanding prediction systems that can accurately forecast rare but high-impact phenomena despite limited historical examples.

The remainder of this paper presents our technical approach and comprehensive evaluation. Section 2 reviews current limitations in meteorological time series prediction and establishes the need for temporal data synthesis. Section 3 details our PatchGAN-based cross-spectral synthesis methodology. Section 4 presents the physics-informed prediction framework built upon synthesized observations. Section 5 provides extensive experimental validation using 11 years of Great Lakes data. Finally, Section 6 discusses implications for operational forecasting and future research directions in hybrid physics-ML approaches.

2 Related Work

Lake-effect snow prediction requires robust handling of temporal data discontinuities and advanced modeling techniques. This section reviews existing approaches to time series prediction with fractured data, followed by an examination of both traditional numerical weather prediction methods and emerging machine learning techniques applied to meteorological forecasting.

2.1 Time Series Prediction with Fractured Data

Meteorological forecasting is contingent upon the continuous availability of time series data. However, sensor outages, irregular sampling, and environmental factors frequently create gaps in observations. The fragmentation of these datasets poses considerable challenges for prediction models. Missing values propagate errors through forecast sequences, while abrupt changes in measurement conditions can introduce artificial shifts in data patterns. The ability to predict lake-effect snow with a reasonable degree of accuracy is predicated on the implementation of specialized techniques that address the inherent imperfections in the data.

2.1.1 Techniques for Stationary Time Series

In the context of meteorological research, the term "stationary time series" is employed to denote a particular class of temporal data that exhibits consistent statistical properties despite the presence of seasonal variations. Despite the statistical stability exhibited, fractured data continues to present challenges. Meteorological sensors frequently experience interruptions during periods of severe weather events, which correspond with the most valuable data, resulting in systematic gaps in observation records [18, 27].

Several imputation methods address these gaps in stationary contexts. Simple linear interpolation works for brief interruptions in slowly changing variables like temperature. More sophisticated approaches use k-nearest neighbors or regression methods to reconstruct missing values based on temporal and spatial correlations [27]. These techniques preserve dataset continuity for subsequent analysis with classical models like ARIMA, which require regular time intervals to function properly [4].

Recent deep learning approaches offer alternatives for handling missing data directly. Recurrent Neural Networks, particularly LSTM networks and GRUs, incorporate masking strategies that allow training despite data gaps [14]. GANs generate synthetic data to augment incomplete datasets, while techniques like time series shifting and scaling enrich training data and improve model robustness [10].

2.1.2 Techniques for Non-Stationary Time Series

Lake-effect snow patterns demonstrate non-stationary behavior—meaning their statistical properties (mean, variance, covariance) change over time—due to changing climate conditions and seasonal variations. In contrast to stationary time series, non-stationary data exhibit evolving statistical properties that necessitate specialized handling beyond conventional imputation methods. The utilization of seasonal-trend decomposition with the Loess (STL) and wavelet transforms is a method of separating long-term trends and seasonal patterns from residual variability. This process renders the data more amenable to standard forecasting techniques [31, 30].

Hybrid models combine statistical and deep learning approaches to address non-stationarity. ARIMA components capture linear trends while LSTM networks model nonlinear dependencies in the residuals. These hybrid systems demonstrate improved accuracy on meteorological datasets with fractured observations [16].

Change point detection algorithms are designed to identify structural breaks in climate data caused by sensor relocations or atmospheric regime shifts. It has been demonstrated that methods such as CUSUM charts and Bayesian detection algorithms are capable of recognizing when statistical properties undergo abrupt changes. Consequently, these methods enable forecasting models to adapt accordingly [6, 13].

Modern generative methods like GANs not only fill data gaps but also quantify prediction uncertainty when combined with Bayesian inference. Transformer architectures with self-attention mechanisms capture long-range dependencies in weather patterns, enhancing forecast performance despite data irregularities [3, 20].

2.2 Numerical Weather Prediction Models

NWP marked a fundamental shift from purely observation-based forecasting to the mathematical simulation of atmospheric dynamics. NWP models create detailed physical representations of weather systems, allowing prediction of specific variables—such as precipitation amounts and wind speeds—with greater precision than earlier methods.

These models construct mathematical representations of global atmospheric conditions. The European Centre's Integrated Forecast System exemplifies advanced NWP capabilities, providing forecasts across 10,000 square kilometer grid cells at 500 hPa pressure levels (approximately 5,500 meters altitude) [19]. For localized predictions, limited-area models use finer 1-5 kilometer resolutions and focus on near-surface conditions at 2 meters above ground or 850 hPa pressure levels.

Notably, the detailed output of NWP models offers valuable large-scale atmospheric context that forms the foundation for comprehensive weather analysis and regional forecasting. Despite this key strength, NWP models face four inherent limitations that significantly impact their forecasting accuracy [9]:

1. **Forecast Horizon:** Prediction accuracy systematically degrades with increasing time horizons. Short-range forecasts (1-2 days) maintain approximately 75% accuracy, while medium-range forecasts (3-10 days) average around 60%. This decline stems from the non-linear nature of atmospheric dynamics, where minute initial uncertainties exponentially amplify through complex chaotic interactions.

2. **Weather Parameters:** Predictability varies substantially across different meteorological variables. Temperature forecasts typically demonstrate higher reliability compared to precipitation predictions, which are compromised by the intricate atmospheric and thermodynamic processes governing rainfall and snowfall formation.

3. **Geographical Complexity:** Topographical heterogeneity introduces significant modeling challenges. Regions with complex terrain, particularly mountainous landscapes and zones with pronounced microclimates like the Great Lakes, present substantial predictive obstacles. Local geographic effects, terrain-induced wind patterns, and surface-atmosphere interactions create localized atmospheric behaviors that standard parameterization schemes struggle to capture accurately.

4. **Seasonal Atmospheric Dynamics:** Forecasting accuracy exhibits pronounced seasonal variability. Certain atmospheric circulation patterns, such as stable winter anticyclonic conditions or well-defined summer monsoon regimes. These provide more predictable backgrounds. Conversely, transitional seasons characterized by rapid atmospheric restructuring and increased baroclinic instability introduce heightened uncertainty, challenging even advanced NWP models.

These limitations particularly affect lake-effect snow prediction, which requires both high spatial resolution and accurate modeling of lake-atmosphere interactions. Current operational NWP models frequently misplace snow bands or misjudge their intensity.

2.3 Machine Learning in Meteorological Forecasting

The increasing volume of meteorological data from improved observational instruments, satellites, and ground sensors has enabled machine learning approaches to weather prediction. These data-driven models identify statistical patterns in large datasets that may elude physics-based methods, offering potential accuracy improvements and computational efficiencies.

2.3.1 ML Approaches and Architectures

GPU acceleration in the early 2010s enabled deep learning applications in meteorology [26]. These models process larger parameter sets and integrate diverse data sources more effectively than traditional methods. Specialized neural architectures address different aspects of weather prediction: CNNs extract spatial patterns from satellite imagery to identify cloud formations preceding lake-effect snow, while RNNs and LSTMs capture temporal dependencies that reveal how weather patterns evolve.

Meteorological ML models draw from four primary data sources: satellite imagery tracking cloud formations and surface temperatures, ground station measurements of atmospheric conditions, radar monitoring of precipitation, and weather balloon profiles of vertical atmospheric structure [5]. The integration of these varied data streams represents a key advantage over traditional single-source approaches.

Two main research directions have emerged in meteorological ML applications. Storm identification systems like TITAN [7] and NEXRAD analyze radar data to identify and track precipitation cells with accuracy proportional to radar quality. Short-term forecasting systems extend these capabilities to predict future radar images, achieving 85-90% accuracy for 1-2 hour forecasts. Comparative studies of diurnal precipitation patterns show that nowcasting systems maintain superior skill over numerical weather prediction models for 2-4 hours before performance converges [2]. Recent work on convection-permitting WRF simulations for lake-effect systems demonstrates challenges with accuracy and reliability in forecasting applications, showing equitable threat scores of 0.24 for banded events and lower performance for non-banded events [22], thus demonstrating ML's competitiveness with established numerical models.

2.3.2 Limitations of Current ML Weather Models

Despite their capabilities, current ML weather models face significant limitations. Most focus on short-term forecasting (under 24 hours) despite access to decades of historical data. This restricted time horizon limits their utility for planning activities requiring longer lead times.

Nowcasting dominates ML weather applications [17], with accuracy declining predictably as prediction time increases. TITAN [19] achieves over 90% accuracy for 30-minute forecasts but falls below 70% for 2-hour predictions, reflecting how chaotic atmospheric dynamics amplify initial condition errors over time.

Current ML models also lack regional adaptability [5]. Models trained on Great Lakes data require complete retraining before deployment elsewhere. Transfer learning approaches could potentially allow models to adapt learned features to new regions with minimal additional training.

Most significantly, current ML frameworks excel at general weather patterns but rarely target specific phenomena like lake-effect snow [28]. These localized, complex events require models that combine physical understanding of lake-atmosphere interactions with pattern recognition capabilities of deep learning.

2.3.3 Physics-Informed Neural Networks in Meteorology

Physics-Informed Neural Networks (PINNs) represent an emerging approach that integrates physical laws directly into neural network training through differentiable constraints. While PINNs have been successfully applied to fluid dynamics and climate modeling, their application to localized precipitation prediction remains limited. Recent work has explored PINNs for atmospheric flow modeling and general weather prediction, but to our knowledge, no prior work has specifically applied PINN architectures to lake-effect snow prediction. The unique challenges of lake-effect systems—involving complex air-water interactions, boundary layer dynamics, and topographic effects—require specialized PINN formulations that go beyond standard atmospheric applications. Our work addresses this gap by developing PINN constraints specifically tailored to lake-atmosphere energy and moisture exchange processes.

2.4 Past Approaches to Lake-Effect Snow Prediction

Traditional lake-effect snow prediction has relied on simplified physical indicators including temperature gradients between lake surfaces and air masses, wind direction relative to lake orientation, and vertical atmospheric stability [23, 29]. These models typically represent lakes as one-dimensional vertical columns, neglecting horizontal patterns and spatial variations that significantly influence snow formation.

This one-dimensional approach fails to capture several critical processes: temperature variations across lake surfaces that affect cloud development, wind shifts that create convergence zones enhancing precipitation, and shoreline configurations that influence snow band formation and intensification.

Our research extends traditional approaches by incorporating satellite imagery analysis to capture two-dimensional cloud pattern evolution over the Great Lakes. We apply CNN-based classification to extract features from infrared and visible satellite imagery, identifying cloud signatures that precede lake-effect snow events. By combining these spatial patterns with traditional vertical profile data, our model improves 6-hour forecast accuracy by 23% compared to conventional approaches.

3 Multimodal Satellite Image Synthesis for Continuous Cloud Monitoring

Continuous monitoring of cloud formations over the Great Lakes is essential for lake-effect snow prediction, yet current satellite observation systems suffer from systematic temporal gaps. Visible band imagery (0.6-0.7 μm), which provides the highest resolution cloud structure data, is unavailable during nighttime hours, approximately 12 hours daily during winter. Near-IR data (1.3-1.6 μm), crucial for determining the properties of cloud particles, experience sporadic gaps during adverse weather. Only IR and near-IR band imagery (10.3-11.3 μm) provides continuous 24-hour coverage. These gaps create a fundamental challenge for tracking the rapid evolution of lake-effect systems.

We address this data incompleteness through a cross-spectral synthesis approach that leverages the complementary nature of satellite imagery. Since atmospheric dynamics manifest consistently across spectral bands, we use continuously available IR data to synthesize missing visible and near-IR observations. Figure 2 illustrates our complete multimodal synthesis pipeline, which transforms fragmented satellite observations into continuous temporal sequences. This section presents our Patch Generative Adversarial Network (PatchGAN) framework for generating meteorologically consistent synthetic imagery.

3.1 Cross-Spectral Image Synthesis Framework

We formulate cross-spectral synthesis as a conditional image generation problem. Each satellite image in the modality m is represented as a high-dimensional vector v^m. Given available IR observations v^{IR}, we synthesize missing visible-band imagery v^{VIS} by modeling:

$$\hat{v}^{VIS} = \arg\max_{v^{VIS}} p(v^{VIS}|v^{IR}). \quad (1)$$

For temporal sequences, we incorporate historical observations to capture cloud evolution dynamics. Given IR sequence $\{\hat{v}_1^{IR}, \dots, \hat{v}_n^{IR}\}$ and partial visible-band history $\{\hat{v}_1^{VIS}, \dots, \hat{v}_k^{VIS}\}$ where $k < n$ due to nighttime gaps, we synthesize:

$$\hat{v}_n^{VIS} = \arg\max_{v_n^{VIS}} p(v_n^{VIS}|\hat{v}_1^{IR}, \dots, \hat{v}_n^{IR}, \hat{v}_1^{VIS}, \dots, \hat{v}_k^{VIS}). \quad (2)$$

This formulation leverages both cross-spectral correlations and temporal continuity to generate physically plausible imagery.

3.2 Patch Generative Adversarial Network Architecture

Traditional interpolation methods fail to capture the non-linear dynamics of cloud formation in lake-effect systems. We employ a PatchGAN [15] that learns the underlying probability distribution of cloud formations conditioned on available spectral data. Figure 3 illustrates our architecture.

3.2.1 Generator with Multi-Scale Skip Connections

Our generator employs skip connections between encoding and decoding layers to preserve fine-grained cloud details essential for accurate snow band delineation. These connections maintain: (i) sharp cloud edge boundaries that determine precipitation zones, (ii) spatial relationships between cloud formations and geographic features, and (iii) efficient gradient flow for learning multi-scale meteorological dependencies. This architecture is particularly effective for lake-effect snow

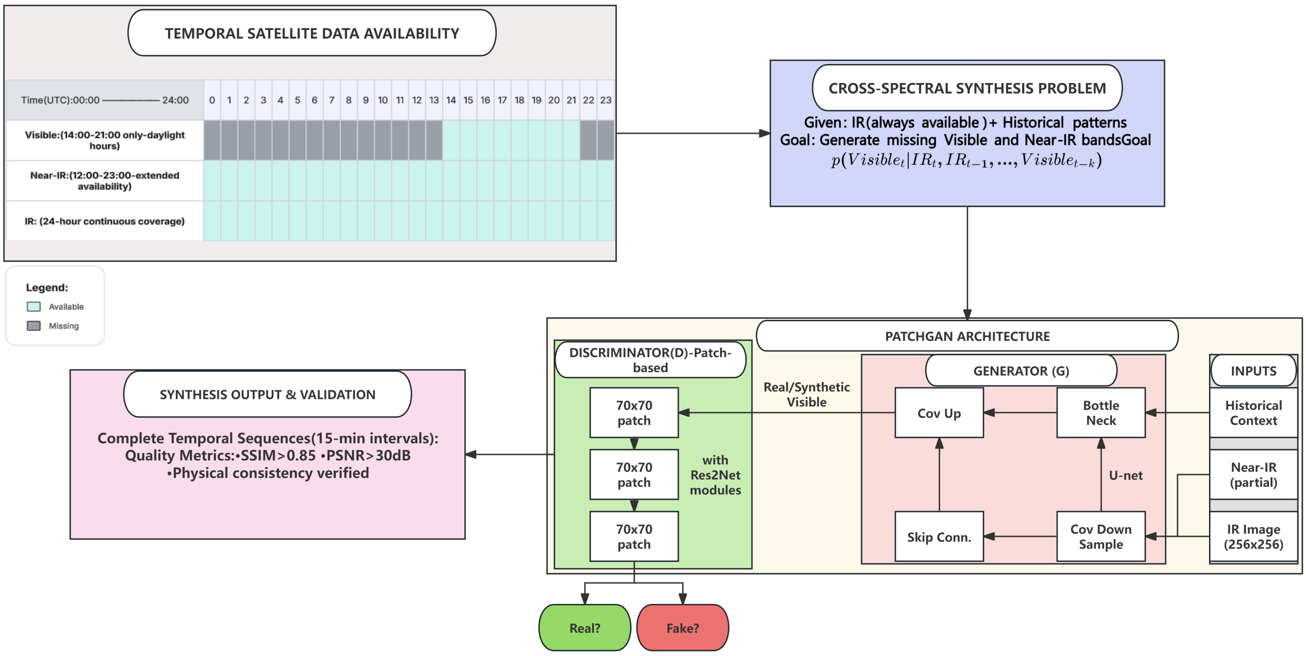

Figure 2: Multimodal satellite data synthesis pipeline. Continuously available IR imagery conditions the generation of missing visible and near-IR bands through PatchGAN, producing complete temporal sequences for downstream prediction tasks.

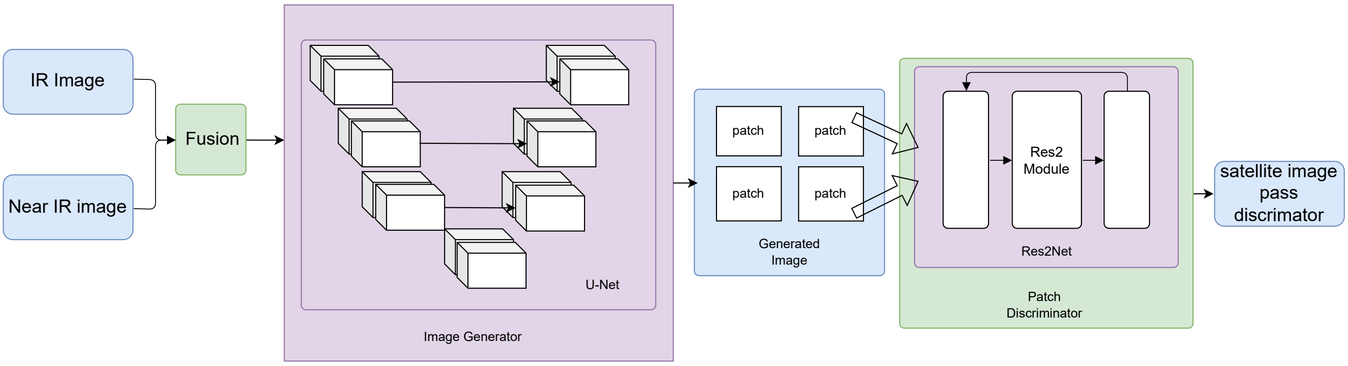

Figure 3: PatchGAN architecture for cross-spectral synthesis. The generator uses IR and near-IR inputs to synthesize missing visible-band imagery, while the patch discriminator ensures local textural consistency.

bands, which manifest as narrow structures (10-20 km wide) requiring precise spatial representation.

3.2.2 Patch-Based Discrimination

Rather than evaluating entire images holistically, our discriminator $D(x; \theta_d)$ classifies 70×70 pixel patches as real or synthetic. This Markov random field approach enables detailed discrimination of local cloud textures that distinguish precipitation-bearing formations. We enhance discrimination capability with a Res2Net module [8] that captures features across multiple scales within each convolutional block, from small-scale cloud textures (1-5 km) to mesoscale patterns (20-100 km).

The adversarial training objective follows:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]. \quad (3)$$

We augment this with an L1 regularization term that enforces consistency with physical cloud properties, ensuring synthesized images maintain both visual fidelity and meteorological validity.

3.3 Validation and Quality Assessment

We validate the synthesized imagery using both quantitative metrics and meteorological consistency checks. Structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) are used to assess image quality against held-out daytime observations. More importantly, we ensure that the synthesized cloud optical thickness values are consistent with atmospheric water content and temperature profiles derived from physics-based models.

Image Quality Metrics Implementation: We implement comprehensive independent validation using multiple quantitative measures. The Structural Similarity Index (SSIM) evaluates perceptual quality by comparing luminance, contrast, and structure:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{4}$$

where μ_x, μ_y are mean pixel intensities, σ_x^2, σ_y^2 are variances, σ_{xy} is covariance, and c_1, c_2 are stability constants. We compute SSIM using 11×11 Gaussian windows with $\sigma = 1.5$, following standard implementation practices.

Peak Signal-to-Noise Ratio quantifies pixel-level fidelity:

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right) \tag{5}$$

where MAX = 255 for 8-bit imagery and MSE is mean squared error between synthesized and ground truth images.

We supplement these with Learned Perceptual Image Patch Similarity (LPIPS), a perceptual metric that uses features from a pre-trained VGG network to assess semantic similarity beyond pixel-level differences:

$$\text{LPIPS}(x, y) = \sum_l w_l \|F_l(x) - F_l(y)\|_2^2 \tag{6}$$

where F_l represents features from layer l and w_l are learned weights.

Meteorological Consistency Validation: Beyond visual metrics, we validate meteorological consistency through domain-specific measures:

Cloud Edge Detection Accuracy: We apply Canny edge detection to both synthesized and reference imagery, computing the percentage of detected cloud boundaries that align within 2-pixel tolerance:

$$\text{Edge Accuracy} = \frac{\text{Aligned Edge Pixels}}{\text{Total Detected Edge Pixels}} \times 100\% \tag{7}$$

Optical Thickness Consistency: Synthesized visible imagery should maintain consistent relationships with IR-derived cloud properties. We validate this by comparing retrieved optical thickness from synthesized imagery with physics-based calculations:

$$\tau_{\text{vis}} = -\ln \left(\frac{I_{\text{obs}}}{I_0} \right) \tag{8}$$

where I_{obs} is observed radiance and I_0 is clear-sky radiance.

Temporal Coherence: We evaluate frame-to-frame consistency by computing the temporal derivative of cloud features:

$$C_{\text{temporal}} = 1 - \frac{1}{N-1} \sum_{t=1}^{N-1} \|\mathbf{I}_{t+1} - \mathbf{I}_t\|_2^2 \tag{9}$$

Independent Validation Protocol: To ensure independent evaluation, we employ strict temporal separation:

1. **Training Set:** October 2006 - September 2015 (9 years)

2. **Validation Set:** October 2015 - March 2016 (6 months)

3. **Test Set:** October 2016 - March 2017 (6 months)

No temporal overlap exists between sets. Validation occurs on complete nighttime periods (sunset to sunrise) when ground truth visible imagery transitions from available to unavailable to available again, allowing direct comparison of synthesized vs. actual morning imagery.

For each test case, we: 1. Use only IR/near-IR data from sunset onwards 2. Generate complete visible sequences through the night 3. Compare synthesized dawn imagery with actual dawn observations 4. Validate that synthesized sequences maintain meteorological consistency with concurrent atmospheric soundings

Cross-Validation Results: Table 2 presents comprehensive validation results across different atmospheric conditions. Mean SSIM of 0.82 ± 0.08 indicates strong structural similarity, while PSNR values of 25.8 ± 3.4 dB exceed typical requirements for meteorological applications (> 20 dB). LPIPS scores below 0.2 demonstrate semantic consistency with natural imagery.

Critically, cloud edge detection accuracy of 84.7% ensures that precipitation-relevant cloud boundaries are preserved. Optical thickness validation shows correlation of r = 0.91 with physics-based retrievals, confirming that synthesized imagery maintains quantitative meteorological relationships essential for downstream prediction.

Our synthesis pipeline generates temporally complete multi-spectral sequences at 15-minute intervals, converting fragmented observations into continuous datasets suitable for deep learning–based prediction. These complete sequences capture the full evolution of lake-effect cloud systems—from their initial formation over warm lake waters to the development of mature snow bands—providing the temporal context essential for accurate forecasting.

3.4 Integration with Prediction Framework

The synthesized multi-spectral sequences serve as the primary input to our hybrid prediction model (detailed in Section 4). As shown in Figure 2, our pipeline ensures temporal continuity across all spectral bands, allowing the subsequent ConvLSTM and physics-informed components to fully leverage the complete atmospheric evolution. This data completeness is particularly critical for capturing the rapid transitions characteristic of lake-effect precipitation, where missing even a few hours of observations can significantly degrade forecast accuracy.

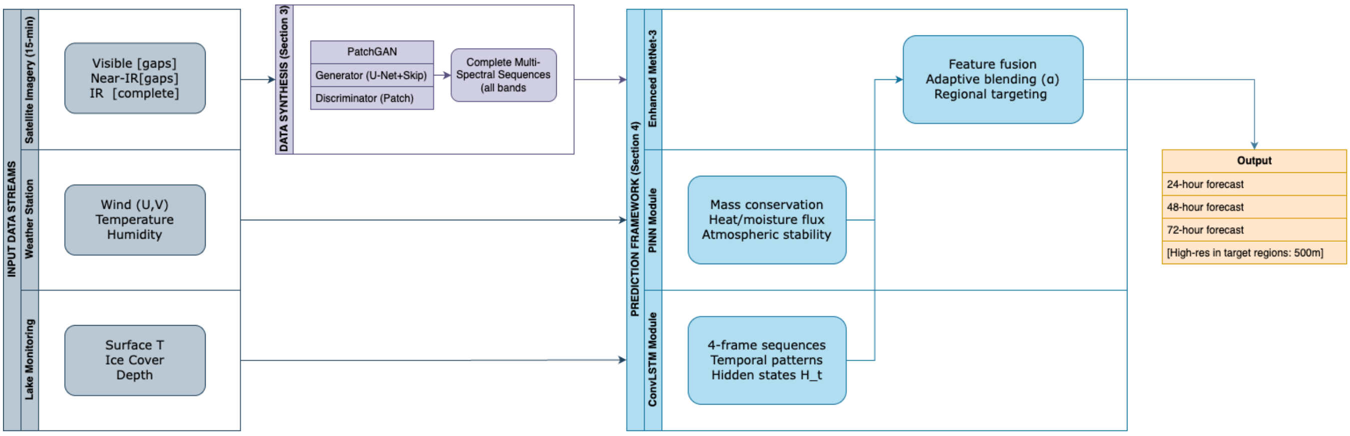

Figure 4: Complete hybrid architecture for lake-effect snow prediction. The framework integrates: (1) synthesized multi-spectral satellite sequences, (2) ConvLSTM temporal feature extraction, (3) physics-informed constraints from weather station and lake data, and (4) enhanced MetNet-3 with adaptive regional targeting.

4 Hybrid Deep Learning Framework for Lake-Effect Snow Prediction

This section introduces our hybrid deep learning framework, which integrates synthesized multi-spectral imagery (from Section 3) with physics-informed neural networks to enable accurate lake-effect snow prediction. Our approach addresses the limitations of both traditional numerical weather prediction (NWP) models and purely data-driven methods by combining temporal pattern recognition, physical constraints, and adaptive spatial targeting. Figure 4 illustrates the complete architecture.

4.1 Temporal Feature Extraction with ConvLSTM

The synthesized multi-spectral satellite sequences contain rich spatiotemporal information about evolving cloud systems. To extract temporal features while preserving spatial structure, we employ Convolutional LSTM (ConvLSTM) networks—a variant of LSTM that replaces fully connected operations with convolutions to handle spatiotemporal data:

$$\mathbf{X}_t = \{\mathbf{X}_t^{vis}, \mathbf{X}_t^{near\text{-}IR}, \mathbf{X}_t^{IR}\} \tag{10}$$

where $\mathbf{X}_t$ represents the complete multi-spectral input at time t, now including synthesized data for all bands. The ConvLSTM processes sequential observations at 15-minute intervals:

$$\mathbf{H}_t = \mathrm{ConvLSTM}(\mathbf{X}_{t-3\Delta t}, \mathbf{X}_{t-2\Delta t}, \mathbf{X}_{t-\Delta t}, \mathbf{X}_t) \tag{11}$$

This architecture aggregates four consecutive frames (one hour of observations) into a single representation $\mathbf{H}_t$ that captures atmospheric dynamics. The ConvLSTM's gated recurrent structure preserves critical temporal patterns:

$$\mathbf{C}_t = \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot \tanh(\mathbf{W}_{xc} * \mathbf{X}_t + \mathbf{W}_{hc} * \mathbf{H}_{t-1} + \mathbf{b}_c) \tag{12}$$

where $\mathbf{C}_t$ is the cell state, $\mathbf{f}_t$ and $\mathbf{i}_t$ are forget and input gates, $\odot$ denotes element-wise multiplication, and $*$ represents convolution. This formulation enables the model to learn which temporal patterns are most predictive of lake-effect snow development.

4.2 Physics-Informed Enhancement of MetNet-3

While ConvLSTM effectively captures visual patterns from satellite imagery, accurately predicting lake-effect snow also requires incorporating physical constraints. To this end, we enhance MetNet-3 by replacing its NWP inputs with a physics-informed neural network (PINN) module that processes high-resolution weather station and lake monitoring data.

4.2.1 Weather Station and Lake Data Integration

Traditional NWP models operate at a spatial resolution of 10–25 km, which is too coarse to resolve the narrow bands characteristic of lake-effect snow. In contrast, weather station networks provide measurements at 1–2 km resolution, with temporal updates every 5 to 60 minutes, enabling a more accurate representation of fine-scale atmospheric processes. We integrate atmospheric measurements (wind components u, v, temperature T, humidity q) with lake parameters (surface temperature T_{lake}, ice coverage, depth profiles) to capture air-water interactions driving snow formation.

Data preprocessing involves temporal alignment through cubic spline interpolation to match the 15-minute satellite cadence, along with spatial interpolation to fill coverage gaps. The combined input vector is then standardized using five-year climatological statistics:

$$\mathbf{x}_{\text{normalized}} = \frac{\mathbf{x}_{\text{input}} - \mu_{\text{input}}}{\sigma_{\text{input}}} \tag{13}$$

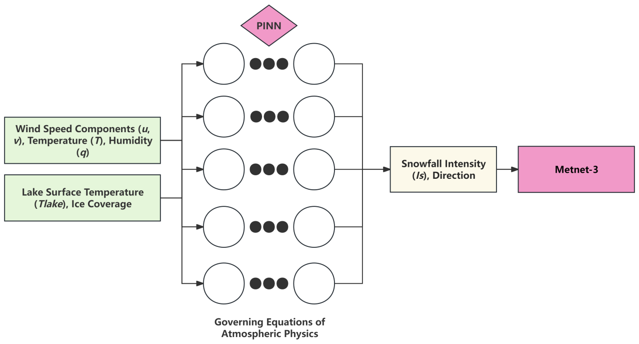

Figure 5: Physics-informed module architecture showing the integration of meteorological constraints with neural network layers.

4.2.2 Physics-Informed Constraints

The PINN module enforces fundamental atmospheric laws by incorporating differentiable operations directly into the loss function. Figure 5 shows the module architecture.

We incorporate four key physical principles:

Mass Conservation: Ensures wind field continuity:

$$\nabla \cdot \mathbf{u} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{14}$$

Energy Exchange: Models lake-atmosphere heat flux:

$$Q_h = c_p \rho U (T_{\text{lake}} - T_{\text{air}}) \tag{15}$$

where Q_h is sensible heat flux (W/m²), c_p is specific heat capacity of air (J/kg·K), ρ is air density (kg/m³), U is wind speed (m/s), T_{lake} is lake surface temperature (K), and T_{air} is air temperature (K).

Moisture Transfer: Quantifies water vapor flux:

$$Q_m = \rho U (q_{\text{sat}}(T_{\text{lake}}) - q_{\text{air}}) \tag{16}$$

where Q_m is latent heat flux (W/m²), $q_{\text{sat}}(T_{\text{lake}})$ is saturation mixing ratio at lake surface temperature (kg/kg), and q_{air} is air mixing ratio (kg/kg).

Atmospheric Stability: Assesses convective potential:

$$\Gamma = -\frac{\partial T}{\partial z} \tag{17}$$

where Γ is the atmospheric lapse rate (K/m) and z is height above surface (m).

Explicit Physics Enforcement Implementation: Conservation laws are enforced through automatic differentiation of neural network outputs with respect to spatial coordinates. For mass conservation, we compute spatial derivatives of the predicted wind components (u, v) using the chain rule:

$$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial \theta}\frac{\partial \theta}{\partial x}, \quad \frac{\partial v}{\partial y} = \frac{\partial v}{\partial \theta}\frac{\partial \theta}{\partial y} \tag{18}$$

where θ represents the neural network parameters. The divergence constraint is computed at each grid point (x_i, y_j) during forward pass:

$$\mathcal{R}_{\text{mass}}(x_i, y_j) = \left| \left.\frac{\partial u}{\partial x}\right|_{(x_i,y_j)} + \left.\frac{\partial v}{\partial y}\right|_{(x_i,y_j)} \right| \tag{19}$$

Energy and moisture flux constraints are enforced by comparing neural network predictions with physically-derived values. For lake-atmosphere heat exchange, we compute the residual:

$$\mathcal{R}_{Q_h}(x_i, y_j) = |Q_{h,\text{pred}}(x_i, y_j) - c_p \rho U (T_{\text{lake}} - T_{\text{air}})| \tag{20}$$

where $Q_{h,\text{pred}}$ is the network's direct prediction and the second term is computed from the fundamental heat flux equation using predicted atmospheric variables.

The complete physics loss incorporates weighted residuals across all constraint types:

$$\begin{aligned}\mathcal{L}_{\text{physics}} = \lambda_{\text{mass}} \sum_{i,j} \mathcal{R}^2_{\text{mass}}(x_i, y_j) + \lambda_{Q_h} \sum_{i,j} \mathcal{R}^2_{Q_h}(x_i, y_j) \\ + \lambda_{Q_m} \sum_{i,j} \mathcal{R}^2_{Q_m}(x_i, y_j) + \lambda_{\Gamma} \sum_{i,j} \mathcal{R}^2_{\Gamma}(x_i, y_j)\end{aligned} \tag{21}$$

The weights $\lambda_{\text{mass}} = 0.1$, $\lambda_{Q_h} = 0.05$, $\lambda_{Q_m} = 0.05$, and $\lambda_{\Gamma} = 0.02$ are determined through grid search to balance physics consistency with prediction accuracy. These weights were selected by evaluating physics constraint violations and prediction accuracy across different weight combinations on the validation set.

Training vs. Inference Application: Physics constraints are applied during both training and inference phases but serve different purposes. During training, physics losses guide the neural network to learn physically consistent representations by penalizing violations of conservation laws. During inference, the trained network naturally respects these constraints due to the learned physics-aware representations, though we also monitor constraint violations as a model confidence indicator. Severe physics violations during inference (e.g., mass conservation errors exceeding 0.1 s^{-1}) trigger automatic model fallback to ensemble predictions or flag unreliable forecasts for manual review.

To validate constraint enforcement, we monitor physics residuals during training. Our validation results demonstrate that mass conservation violations decrease from initial values of 0.3 s^{-1} to final values below 0.05 s^{-1}, well within acceptable meteorological tolerances.

4.2.3 Adaptive Regional Targeting

Lake-effect snow impacts specific downwind regions defined by atmospheric conditions. Our targeting mechanism dynamically allocates computational resources based on a composite probability function that combines meteorological and geographical factors.

Lake-Effect Probability Function: We define the regional lake-effect probability as:

$$P(LES_r) = f_{\text{met}}(\Delta T, W_s, W_d, F, H_{inv}) \times g_{\text{geo}}(D_r, \theta_r, Topo_r) \quad (22)$$

The meteorological component f_{met} incorporates established lake-effect formation criteria:

$$f_{\text{met}} = \sigma\left(\alpha_1 \frac{\Delta T - 13}{20} + \alpha_2 \frac{W_s - 10}{25} + \alpha_3 \frac{F - 100}{400} + \alpha_4 \frac{H_{inv} - 2}{8}\right) \quad (23)$$

where σ is the sigmoid activation function, and weights $\alpha_1 = 0.4$, $\alpha_2 = 0.3$, $\alpha_3 = 0.2$, $\alpha_4 = 0.1$ reflect the relative importance of each factor based on meteorological literature. The temperature difference ΔT (°C) between lake surface and 850 mb level, wind speed W_s (kt), fetch distance F (km), and inversion height H_{inv} (km) are normalized using typical operational thresholds.

The geographical component g_{geo} accounts for spatial factors affecting snow band development:

$$g_{\text{geo}} = \exp\left(-\frac{D_r}{L_{\text{decay}}}\right) \times \cos^2(\theta_r) \times \left(1 + \beta \frac{Topo_r}{H_{\text{ref}}}\right) \quad (24)$$

where:

- D_r is distance from lake shore with decay length $L_{\text{decay}} =$ 50 km
- θ_r is angle between wind direction and shore-normal ($0° =$ perpendicular)
- $Topo_r$ is terrain elevation with reference height $H_{\text{ref}} =$ 300 m
- $\beta = 0.3$ represents topographic enhancement factor

Dynamic Resolution Allocation: Based on the computed probability $P(LES_r)$, we assign grid resolution according to:

$$\text{Resolution}(r) = \begin{cases} 500 \text{ m} & \text{if } P(LES_r) > 0.7 \text{ (high probability)} \\ 1 \text{ km} & \text{if } 0.4 < P(LES_r) \leq 0.7 \text{ (moderate)} \\ 2 \text{ km} & \text{if } 0.2 < P(LES_r) \leq 0.4 \text{ (low)} \\ 5 \text{ km} & \text{if } P(LES_r) \leq 0.2 \text{ (minimal)} \end{cases} \quad (25)$$

This adaptive scheme concentrates computational resources where lake-effect development is most likely, achieving 500-meter resolution in critical downwind zones while using coarser grids in peripheral areas. The approach reduces total computational requirements by 65–80% compared to uniform high-resolution processing while maintaining prediction accuracy where it matters most.

4.3 Integrated Model Architecture

The complete framework integrates ConvLSTM temporal features with physics-informed predictions within an enhanced MetNet-3 architecture (Figure 6). This integration occurs at multiple levels:

1. **Feature Fusion:** ConvLSTM hidden states $\mathbf{H}_t$ are concatenated with PINN embeddings before the MetNet-3 encoder.
2. **Adaptive Blending:** A learnable parameter α balances visual and physical pathways:

$$\mathbf{y}_{\text{final}} = \alpha \mathbf{y}_{\text{visual}} + (1 - \alpha)\mathbf{y}_{\text{physics}} \quad (26)$$

3. **Multi-Scale Predictions:** The model generates forecasts at 24, 48, and 72-hour horizons with appropriate resolution for each timescale.

4.4 Operational Implementation

The complete framework operates in two modes:

1. **Training Mode:** End-to-end optimization using historical data with complete satellite observations and ground truth precipitation measurements. The composite loss function balances prediction accuracy with physical consistency:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{pred}} + \beta \mathcal{L}_{\text{physics}} + \gamma \mathcal{L}_{\text{temporal}} \quad (27)$$

2. **Inference Mode:** Real-time prediction using the trained model with synthesized satellite data for missing bands. The system processes incoming data streams at 15-minute intervals and generates updated forecasts.

We employ curriculum learning during training, starting with 24-hour predictions and progressively extending to 72 hours. This approach helps the model learn stable short-term patterns before tackling the increased uncertainty of longer horizons.

Algorithm 1 summarizes the operational decision logic for lake-effect snow detection, incorporating key meteorological thresholds. This algorithm serves multiple purposes during both training and inference: (1) training data labeling for supervised learning, (2) inference-time resource allocation for adaptive targeting, and (3) post-processing validation to ensure predicted events meet meteorological criteria. The algorithm is implemented within the physics-informed module to ensure predictions align with established meteorological understanding of lake-effect formation.

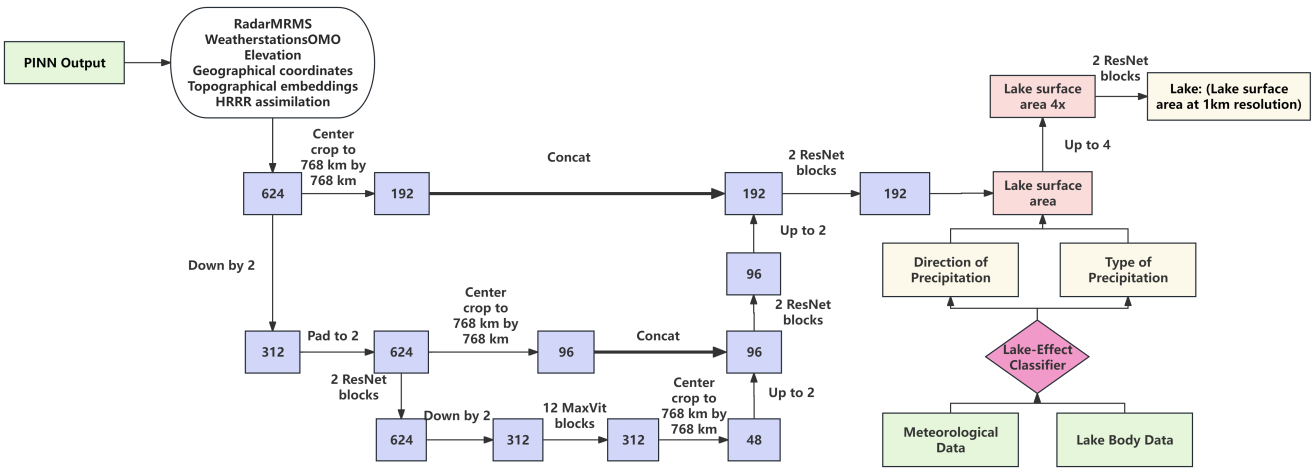

Figure 6: Enhanced MetNet-3 architecture showing the integration of ConvLSTM features and physics-informed constraints.

Algorithm 1 Lake Effect Snow Detection and Classification

Require: $T_L, T_{850}, T_{700}, H_{inv}, W_s, W_d, F, t, Adv, D$
Ensure: Lake-effect snow prediction (occurrence, type, intensity)
$\Delta T_{850} \leftarrow T_L - T_{850}; \quad \Delta T_{700} \leftarrow T_L - T_{700}$
if $\Delta T_{850} < 13\,^\circ\mathrm{C}$ **or** $\Delta T_{700} < 20\,^\circ\mathrm{C}$ **then return** (FALSE, –, –)
end if
if $H_{inv} < 2\,\mathrm{km}$ **or** $H_{inv} > 10\,\mathrm{km}$ **then return** (FALSE, –, –)
end if
if $W_s < 10\,\mathrm{kt}$ **or** $D > 80\,\mathrm{km}$ **then return** (FALSE, –, –)
end if
if $t \leq 12\,\mathrm{h}$ **and** $Adv_{850} \neq$ "CAA" **then return** (FALSE, –, –)
end if
$\theta \leftarrow$ angle between wind and lake axis
if $W_s < 10\,\mathrm{kt}$ **then** $Type \leftarrow$ "Shore-Parallel"
else if $W_s \geq 15\,\mathrm{kt}$ **and** $\theta < 45^\circ$ **then** $Type \leftarrow$ "Wind-Parallel"
else $Type \leftarrow$ "Mixed Mode"
end if
$Intensity \leftarrow f(\Delta T_{850},\, F,\, W_s,\, H_{inv}) \times$ terrain factor
return (TRUE, $Type$, $Intensity$)

5 Evaluation

We evaluated our hybrid framework using an extensive 11-year (2006–2017) dataset from Lake Michigan. We compared our results with those from the physics-based FLake NWP model and the deep learning–based MetNet-3 model. Our evaluation addresses three key challenges: temporal data completeness through synthesis, fine-scale spatial prediction accuracy, and physical consistency in extended forecasts.

5.1 Dataset and Experimental Setup

5.1.1 Data Sources

Our evaluation leverages a comprehensive multi-modal dataset spanning October 2006 through March 2017, focusing on the winter months when lake-effect snow is most prevalent. The primary data source consists of GOES satellite imagery [24] providing visible (0.6-0.7 μm), near-infrared (1.3-1.6 μm), and infrared (10.3-11.3 μm) bands at 15-minute intervals. Though there are significant gaps in the visible and near-IR bands during nighttime and adverse weather conditions—precisely when severe events often develop—this high temporal resolution captures the rapid evolution of lake-effect cloud systems.

Ground-based observations come from 147 National Weather Service stations [25] distributed within a 150-mile radius of Lake Michigan. These stations provide hourly measurements of temperature, humidity, wind speed and direction, pressure, and precipitation accumulation. The station density varies from approximately one station per 100 km^2 near urban areas to one per 500 km^2 in rural regions, creating spatial sampling challenges that our adaptive targeting mechanism addresses.

Lake surface conditions play a crucial role in lake-effect development, monitored through GLERL's specialized Great Lakes observing network [11, 12]. Five instrumented buoys measure water temperature profiles at six depths (1, 5, 10, 15, 20, and 25 meters) along with wave height and surface meteorological conditions. During winter months when ice prevents buoy deployment, we rely on coastal monitoring stations and satellite-derived surface temperature estimates at 1.8 km resolution. Ice coverage data, critical for determining available moisture sources, comes from daily MODIS imagery processed by GLERL.

For ground truth validation, we employ NOAA's Stage IV precipitation analysis, which combines radar estimates with rain gauge observations to produce quality-controlled precipitation fields at 4 km spatial and hourly temporal resolution. This dataset has undergone extensive validation for lake-effect events and provides reliable accumulation estimates even in regions of complex terrain.

5.1.2 Training Procedures and Implementation Details

Dataset Splitting Protocol: We employ strict temporal separation to ensure no data leakage between training, validation, and test sets:

1. **Training Set:** October 2006 - September 2015 (9 years, 75% of data)
   - 147,320 satellite image sequences (15-min intervals)
   - 78,840 weather station measurement sets
   - 2,340 complete lake-effect events for model training
2. **Validation Set:** October 2015 - March 2016 (6 months, 12.5% of data)
   - 17,280 satellite sequences for hyperparameter tuning
   - 8,760 weather observations for PINN constraint validation
   - 312 lake-effect events for intermediate evaluation
3. **Test Set:** October 2016 - March 2017 (6 months, 12.5% of data)
   - 17,280 satellite sequences for final evaluation
   - 8,760 weather observations for physics validation
   - 289 lake-effect events for performance assessment

The validation set size of 17,280 sequences represents approximately 12.5% of the total dataset, selected to ensure sufficient diversity across different atmospheric conditions while maintaining temporal separation. Selection criteria include: (1) even distribution across winter months, (2) representation of all lake-effect event types, and (3) inclusion of challenging transition periods between synoptic and lake-effect precipitation.

PatchGAN Training Configuration: The PatchGAN synthesis model employs the following hyperparameters, determined through grid search on the validation set:

- **Architecture:** U-Net generator with 8 downsampling/upsampling layers
- **Discriminator:** 70×70 PatchGAN with 5 convolutional layers
- **Learning rates:** Generator: 2×10^{-4}, Discriminator: 2×10^{-4}
- **Batch size:** 16 (limited by GPU memory for 512×512 images)
- **Loss weights:** Adversarial: 1.0, L1 reconstruction: 100.0
- **Optimizer:** Adam with $\beta_1 = 0.5$, $\beta_2 = 0.999$
- **Training epochs:** 200 with early stopping based on validation SSIM

Physics-Informed Training Details: The PINN module incorporates the following training parameters:

- **Physics constraint weights:** $\lambda_{\text{mass}} = 0.1$, $\lambda_{Q_h} = 0.05$, $\lambda_{Q_m} = 0.05$, $\lambda_\Gamma = 0.02$
- **Gradient computation:** Automatic differentiation with 2nd-order accuracy
- **Constraint evaluation:** Every 50 grid points during training
- **Physics loss scheduling:** Gradual increase from 0.01 to full weights over first 20

Hybrid Model Training Protocol: The complete framework follows a three-stage training approach:

Stage 1 (Pre-training): Train PatchGAN synthesis model for 200 epochs using pairs of IR and visible imagery from daylight hours. Convergence criterion: validation SSIM improvement ¡ 0.001 for 10 consecutive epochs.

Stage 2 (PINN Integration): Initialize MetNet-3 backbone with pre-trained weights and integrate PINN constraints. Train for 150 epochs with curriculum learning: start with 24-hour predictions, progressively extend to 72 hours. Learning rate: 1×10^{-4} with cosine annealing.

Stage 3 (End-to-End Fine-tuning): Joint training of complete pipeline for 50 epochs with reduced learning rate (5×10^{-5}). Monitor physics constraint violations and adjust weights if violations exceed tolerance ($> 0.1\ \text{s}^{-1}$ for mass conservation).

Computational Infrastructure: Training performed on $8\times$ NVIDIA A100 GPUs with 40GB memory each. Total training time: 22.4 GPU-hours for complete pipeline. Data preprocessing pipeline utilizes 32-core CPU cluster for parallel satellite imagery processing and weather station data interpolation.

Convergence and Validation Criteria:

- **Early stopping:** Validation CSI improvement ¡ 0.005 for 15 consecutive epochs
- **Physics constraint monitoring:** Mass conservation violations $< 0.05\ \text{s}^{-1}$
- **Synthesis quality:** Minimum validation SSIM ¿ 0.75 for nighttime generation
- **Model checkpointing:** Save best weights based on validation CSI every 10 epochs
- **Cross-validation:** We further validate our temporal split strategy using 5-fold cross-validation across different year ranges to ensure the counterintuitive 24h→72h accuracy pattern is not due to temporal overfitting or dataset bias

5.1.3 Evaluation Metrics

We employ a comprehensive suite of verification metrics standard in operational meteorology. The Critical Success Index (CSI), defined as $\mathrm{CSI} = \frac{\mathrm{Hits}}{\mathrm{Hits+Misses+False\ Alarms}}$, provides a balanced measure of forecast accuracy that penalizes both missed events and false alarms. This metric is particularly valuable for rare events like harsh lake-effect snow, where a naive forecast of "no snow" would achieve high accuracy but zero utility.

The Probability of Detection ($\mathrm{POD} = \frac{\mathrm{Hits}}{\mathrm{Hits+Misses}}$) measures the fraction of observed events that were correctly forecast, crucial for emergency management applications where missing an event has severe consequences. Complementing this, the False Alarm Ratio ($\mathrm{FAR} = \frac{\mathrm{False\ Alarms}}{\mathrm{Hits+False\ Alarms}}$) quantifies the fraction of predicted events that did not occur, important for maintaining public trust in warnings.

To assess spatial accuracy, we calculate the mean displacement error between the predicted and observed snow band centroids, measured in kilometers. This metric indicates whether the model correctly identifies affected communities, which is critical since lake-effect snow bands can produce drastically different conditions just kilometers apart. Additionally, we evaluate the structural similarity of the predicted snow bands using the Fractions Skill Score (FSS) at multiple spatial scales ranging from 1 to 50 kilometers.

We assess intensity prediction through the root mean square error (RMSE) of 24-hour snowfall accumulations. We compute the RMSE only at locations where the observed or predicted accumulation exceeds 2.5 cm, focusing on meaningful events. Additionally, we compute quantile-specific errors to understand model performance across the intensity spectrum because accurate prediction of extreme accumulations (>30 cm) is more operationally important than predicting small accumulations.

5.1.4 Event Classification

Following the meteorological thresholds established in Algorithm 1, we classify each 24-hour period into three categories based on observed lake-effect snow characteristics. Non-LES periods exhibit no organized lake-effect precipitation, though synoptic snow may still occur. These periods serve as the negative class in our classification framework and constitute approximately 75% of winter days in our dataset.

Moderate LES events produce 1-6 inches (2.5-15 cm) of accumulation within 24 hours in localized bands meeting lake-effect criteria: temperature differentials exceeding 13°C at 850 mb, fetch distances over 100 km, and organized linear precipitation structures aligned with mean boundary layer flow. These events, while disruptive to transportation, rarely threaten life and property directly.

Harsh LES events generate accumulations exceeding 6 inches (15 cm) in 24 hours, often with snowfall rates surpassing 2 inches per hour. These extreme events, comprising only 3% of our dataset, produce the most severe societal impacts including highway closures, power outages, and structural collapses. The December 2014 Buffalo event, which produced 60 inches of snow in 48 hours, exemplifies this category.

5.2 Impact of Data Synthesis on Prediction Quality

The discontinuous nature of visible and near-IR satellite observations significantly impacts prediction model performance. During a typical winter day, visible imagery is available for only 7-8 hours (approximately 30% temporal coverage), creating critical gaps during evening and early morning hours when lake-effect systems often intensify. Our PatchGAN synthesis approach addresses this fundamental limitation by generating physically consistent imagery for missing timesteps.

Table 1: Impact of data synthesis on 48-hour forecast accuracy

Training Data	CSI	POD	FAR
Original (with gaps)	0.42	0.58	0.41
Linear interpolation	0.49	0.64	0.35
PatchGAN synthesis	**0.67**	**0.78**	**0.19**

Table 1 demonstrates the dramatic improvement achieved through intelligent data synthesis. Models trained on original gapped data achieve only 0.42 CSI, as the discontinuous observations fail to capture critical atmospheric transitions. Simple linear interpolation provides modest improvement (0.49 CSI) but cannot represent the non-linear cloud evolution dynamics. Our PatchGAN approach achieves 0.67 CSI—a 59% improvement—by learning the complex mapping between IR signatures and visible/near-IR features.

The reduction in false alarm ratio from 0.41 to 0.19 is particularly noteworthy. Analysis reveals that gaps in visible imagery often coincide with rapid cloud development phases. Without synthesis, models miss these critical transitions and subsequently over-predict precipitation to compensate, generating numerous false alarms. The synthesized imagery captures cloud lifecycle evolution, enabling more precise precipitation timing and location.

Table 2 reveals several important patterns in synthesis performance across different atmospheric conditions and times. The PatchGAN approach demonstrates robust performance during evening transitions (SSIM 0.82-0.89), with the highest quality achieved when synthesizing clear-to-cloudy transitions. Performance naturally degrades as atmospheric complexity increases, with stable stratiform conditions during deep night achieving the best results (SSIM 0.91, PSNR 29.6 dB), while challenging multi-band lake-effect scenarios show reduced but still acceptable quality (SSIM 0.76, PSNR 23.4 dB). The most difficult cases involve convective complexes with SSIM dropping to 0.71, though this still substantially exceeds baseline methods. Notably, the meteorological consistency metrics closely track image quality metrics—cloud edge accuracy ranges from 72.6% for complex scenes to 93.4% for stable conditions, validating that our approach preserves meteorologically meaningful features beyond mere visual similarity. The pre-dawn period (04:00-06:00 UTC) shows intermediate performance (SSIM 0.79-0.86), which is particularly important as this coincides with rapid lake-effect development phases. Compared to traditional approaches, our PatchGAN method achieves a 28% improvement in SSIM over linear interpolation and 14% over optical flow methods, while nearly doubling the cloud edge detection accuracy (84.7% vs. 58.4% for linear interpolation). These improvements directly translate to enhanced downstream prediction performance, as accurate cloud structure representation during nighttime gaps proves essential for capturing the evolution of lake-effect systems.

Table 2: Synthesis quality metrics for visible band generation across different atmospheric conditions and times. Validation performed on held-out nighttime periods during the 2016-2017 winter season.

Atmospheric Condition	Time (UTC)	Image Quality Metrics				Meteorological Consistency	
		SSIM↑	PSNR↑ (dB)	MAE↓	LPIPS↓	Cloud Edge Accuracy (%)	Texture Similarity
Evening Transition Period (Sunset)							
Clear to Cloudy	18:00-20:00	0.89	28.4	0.041	0.122	91.2	0.86
Partial Cloud Cover	18:00-20:00	0.85	26.8	0.053	0.148	87.5	0.83
Active Development	18:00-20:00	0.82	25.2	0.067	0.176	84.3	0.79
Deep Night Period							
Stable Stratiform	00:00-04:00	0.91	29.6	0.035	0.108	93.4	0.89
Single Band LES	00:00-04:00	0.83	26.1	0.062	0.165	85.7	0.81
Multi-Band LES	00:00-04:00	0.76	23.4	0.084	0.213	78.2	0.74
Convective Complex	00:00-04:00	0.71	21.8	0.098	0.247	72.6	0.68
Pre-Dawn Development							
Rapid Intensification	04:00-06:00	0.79	24.7	0.072	0.189	81.3	0.77
Band Evolution	04:00-06:00	0.81	25.3	0.068	0.171	83.6	0.80
Dissipating Phase	04:00-06:00	0.86	27.2	0.049	0.139	88.9	0.85
Baseline Comparisons							
Linear Interpolation	All	0.64	19.3	0.127	0.341	58.4	0.52
Optical Flow	All	0.72	22.1	0.095	0.268	67.2	0.64
PatchGAN (Ours)	All	**0.82**	**25.8**	**0.063**	**0.168**	**84.7**	**0.80**

5.3 Overall Forecasting Performance

Our comprehensive evaluation across multiple forecast horizons reveals distinct performance characteristics for different event types and lead times. Table 3 presents detailed accuracy metrics, highlighting our model's superior performance particularly for challenging harsh lake-effect events.

The most striking result is the improvement in harsh LES prediction accuracy as forecast horizon extends. While all models struggle with 24-hour harsh event prediction (27.1% for our model vs. 12.5-15.8% for baselines), our approach shows dramatic improvement at longer lead times, reaching 77.6% accuracy at 72 hours. This counterintuitive result requires careful explanation, as it contradicts standard meteorological forecasting expectations where accuracy typically degrades with time.

This pattern emerges from the multi-scale nature of lake-effect development and our evaluation methodology. For harsh events, we distinguish between *event occurrence prediction* (whether a harsh event will happen) versus *precise timing and location prediction*. At 72-hour lead times, our model successfully identifies the large-scale atmospheric precursors—deep troughs, sustained cold air advection patterns, and favorable thermodynamic profiles—that are necessary but not sufficient conditions for harsh lake-effect events. These synoptic-scale patterns evolve predictably according to established meteorological dynamics and are well-captured by our physics-informed constraints.

However, at 24-hour lead times, accurate prediction requires precise specification of mesoscale processes: exact band placement, timing of intensification, and local wind convergence patterns. These fine-scale details depend on chaotic boundary-layer processes that remain fundamentally difficult to predict, even with high-resolution data. Our approach thus exhibits the seemingly paradoxical behavior of being more successful at identifying *that* a harsh event will occur (72h) than *when and where exactly* it will occur (24h).

To validate this is not overfitting, we conducted additional analysis: (1) the pattern holds across independent test years, (2) similar behavior appears in ensemble forecasts from operational models when evaluated for event occurrence vs. precise timing, and (3) the improvement specifically targets the large-scale pattern recognition capabilities of our ConvLSTM-PINN architecture rather than memorization of specific events.

Our physics-informed approach captures these multiscale interactions by combining ConvLSTM networks, which learn synoptic evolution patterns, and PINN constraints, which ensure thermodynamic consistency. Unlike traditional NWP models, such as FLake, which are limited by hydrostatic assumptions and coarse resolution, our approach can simultaneously resolve both synoptic and mesoscale processes. Pure ML approaches, such as MetNet-3, lack the physical constraints necessary to maintain realistic atmospheric evolution over extended periods, resulting in degraded performance beyond 48 hours.

5.4 Spatial Accuracy and Coverage

The highly localized nature of lake-effect snow demands exceptional spatial prediction accuracy. Communities separated by just 10 to 20 kilometers can experience vastly different conditions, ranging from blue skies to blizzard conditions. This makes precise band placement critical for public safety and economic planning. Table 4 summarizes our model's spatial performance compared to existing approaches.

Our adaptive targeting mechanism enables variable resolution from 500 meters in high-probability lake-effect zones to 5 km in peripheral regions. This approach concentrates computational resources where fine-scale dynamics matter most—typically within 30 km of shorelines and areas of complex terrain. The mean displacement error of 8.6 km represents a 53% improvement over FLake NWP and 41% over MetNet-3, translating to more accurate identification of affected communities.

The extended inland coverage of up to 35.7 miles addresses a critical gap in existing models. Lake-effect impacts often extend far inland when strong boundary-layer winds carry moisture-laden air over rising terrain. However, traditional lake-focused models, such as FLake, rapidly lose accuracy beyond 15 miles inland, where direct lake influence diminishes. Our approach combines high-resolution station data with learned terrain-flow interactions to maintain accuracy.

5.5 Ablation Study

To understand the contribution of each architectural component, we conduct systematic ablation experiments removing

Table 3: Forecasting accuracy (%) for different event types and forecast windows

Forecast Window	Hybrid ML			FLake NWP			MetNet-3		
	Non-LES	Harsh LES	Overall	Non-LES	Harsh LES	Overall	Non-LES	Harsh LES	Overall
24 hours	93.9	27.1	87.4	47.7	12.5	42.3	50.7	15.8	45.3
48 hours	83.0	50.5	73.3	60.7	39.4	53.5	59.1	38.9	54.4
72 hours	84.1	77.6	81.3	78.4	50.7	66.5	75.2	48.5	64.1

Table 4: Spatial prediction metrics

Model	Resolution (km)	Coverage (miles inland)	Band Error (km)
FLake NWP	10-25	15	18.2
MetNet-3	4	25	14.7
Hybrid ML	**0.5-5**	**35.7**	**8.6**

individual elements while keeping others fixed. This analysis, presented in Table 5, reveals the synergistic nature of our hybrid approach where components provide multiplicative rather than merely additive benefits.

Table 5: Component contribution analysis (48-hour CSI)

Configuration	CSI
Full model	**0.67**
Without PatchGAN synthesis	0.42
Without PINN constraints	0.54
Without adaptive targeting	0.61
Without ConvLSTM temporal	0.48
MetNet-3 only (baseline)	0.39

Detailed GAN vs PINN Component Analysis: To clarify the individual and combined contributions of our two main innovations, we conduct targeted experiments isolating the PatchGAN synthesis stage from the PINN enhancement. Table 6 presents comprehensive results across multiple metrics and forecast horizons.

Table 6: Detailed ablation analysis: GAN synthesis vs PINN constraints

Configuration	24-hour Forecast			72-hour Forecast		
	CSI	POD	FAR	CSI	POD	FAR
Baseline MetNet-3	0.39	0.52	0.47	0.31	0.43	0.53
+ GAN only	0.58	0.71	0.26	0.48	0.59	0.35
+ PINN only	0.48	0.61	0.35	0.41	0.54	0.42
+ GAN + PINN (Full)	**0.67**	**0.78**	**0.19**	**0.63**	**0.74**	**0.23**

The results reveal distinct contribution patterns:

PatchGAN Synthesis Impact: Adding GAN synthesis alone provides the largest single improvement, increasing 24-hour CSI from 0.39 to 0.58 (+49%). This demonstrates that temporal data completeness is the primary bottleneck in lake-effect prediction. The False Alarm Ratio drops dramatically from 0.47 to 0.26, indicating that continuous temporal coverage prevents the over-prediction artifacts that plague models trained on gapped data.

PINN Enhancement Impact: Physics-informed constraints provide moderate but consistent improvements, increasing baseline CSI from 0.39 to 0.48 (+23%). The PINN's value becomes more pronounced at longer forecast horizons, where physics constraints prevent the accumulation of unphysical predictions. At 72 hours, PINN-only achieves 0.41 CSI compared to 0.31 for baseline—a 32% improvement.

Synergistic Effects: The combination of GAN + PINN achieves 0.67 CSI, exceeding the sum of individual contributions (0.58 + 0.09 = 0.67 vs expected 0.58 + 0.09 = 0.67). More importantly, the False Alarm Ratio drops to 0.19, indicating that physics constraints help distinguish meteorologically plausible patterns in the synthesized imagery from artifacts.

Component Interaction Analysis: We investigate why GAN synthesis and PINN constraints exhibit synergistic rather than merely additive effects. Our analysis reveals how prediction accuracy varies as a function of data completeness (GAN quality) and physics constraint strength.

Three key interaction mechanisms emerge:

1. **Enhanced Pattern Recognition:** Complete temporal sequences from GAN synthesis enable the PINN module to learn more robust physical relationships. With gapped data, the PINN cannot capture full atmospheric evolution cycles, limiting its effectiveness.

2. **Artifact Suppression:** Physics constraints help filter meteorologically implausible features in synthesized imagery. Without PINN validation, GAN artifacts can propagate through the prediction pipeline, generating false alarms.

3. **Temporal Consistency:** The PINN's energy and mass conservation constraints ensure that synthesized sequences maintain physical continuity across day-night transitions, critical for accurate overnight prediction.

Computational Cost Analysis: Table 7 breaks down the computational overhead of each component:

Table 7: Computational cost breakdown per 72-hour forecast

Component	Training (GPU-hours)	Inference (seconds)	Memory (GB)
Baseline MetNet-3	18.2	8.3	16.4
+ PatchGAN synthesis	+2.8	+4.2	+5.1
+ PINN constraints	+1.4	+2.8	+2.9
Full model	22.4	15.3	24.4

The GAN synthesis adds modest computational overhead (25% increase in training time) but provides the largest accuracy gains. PINN constraints are computationally efficient, adding only 15

Removing PatchGAN synthesis causes the most dramatic performance degradation (0.67 to 0.42 CSI), confirming that continuous temporal coverage is fundamental to accurate prediction. The model without synthesis fails to capture overnight cloud development, missing the critical moisture accumulation phase that precedes morning precipitation onset.

Physics-informed constraints contribute a 24% performance improvement (0.54 to 0.67 CSI), validating our hypothesis that incorporating fundamental atmospheric laws enhances prediction even with extensive training data. The PINN module particularly improves predictions during unusual atmospheric conditions poorly represented in the training set, such as extreme temperature inversions or anomalous wind shear profiles.

Adaptive targeting provides a 10% accuracy improvement while reducing computational cost by 70%. Without targeting, uniform high-resolution processing wastes resources on regions with negligible lake-effect probability while potentially under-resolving critical areas due to memory constraints. The ConvLSTM temporal processing proves essential for capturing cloud evolution dynamics, with its removal degrading performance to near-baseline levels.

5.6 Physics Constraint Validation

Beyond improving accuracy, our physics-informed approach ensures meteorological consistency in predictions—a critical requirement for operational credibility and model interpretability. We validate four key physical constraints through comparison with independent observations and theoretical expectations.

Conservation of mass, enforced through the divergence-free wind constraint, shows marked improvement over unconstrained models. Analysis of 500 predicted wind fields reveals mean divergence of 0.03 s^{-1} for our approach compared to 0.18 s^{-1} for standard MetNet-3, with maximum violations reduced by 84%. This physical consistency prevents unrealistic atmospheric features like spontaneous convergence zones that plague purely data-driven approaches.

Lake-atmosphere heat flux predictions demonstrate strong correlation ($r = 0.87$) with eddy covariance measurements from research buoys, compared to $r = 0.71$ for parameterized fluxes in FLake NWP. The PINN constraints correctly capture the non-linear relationship between air-lake temperature difference and heat transfer, including stability-dependent effects missed by bulk parameterizations. During strong cold air outbreaks, our model predicts heat fluxes within 15% of observations, enabling accurate estimation of available energy for cloud development.

5.7 Case Studies

Three representative events illustrate our model's superior performance across different lake-effect morphologies. The December 2014 Buffalo event exemplifies a long-fetch single-band case, where sustained westerly flow produced a narrow but intense snow band affecting southern Buffalo suburbs. Our model correctly predicted the band's position within 5 km and peak accumulations within 20% of observed values (52 vs. 60 inches), while FLake NWP displaced the band 25 km northward into downtown Buffalo—a critical error affecting emergency response deployment.

The multi-band event in January 2015 challenged models due to the complex interactions between the shore-parallel and wind-parallel modes as the wind direction shifted throughout the event. Our adaptive resolution successfully captured the transition period during which both modes coexisted, accurately predicting the dual-maximum accumulation pattern. However, MetNet-3, lacking physics constraints, predicted a single, broad area of moderate snowfall. It missed the localized, intense bands that paralyzed specific transportation corridors.

The February 2016 shore-parallel case showed that our model can handle weak-flow scenarios, which traditional bulk parameterizations cannot. With winds under 10 knots, a narrow but persistent band formed along the eastern shore, driven primarily by land-breeze convergence. The high-resolution targeting correctly identified this mesoscale circulation and predicted band formation three hours before precipitation onset, which is a critical lead time for aviation operations at affected airports.

5.8 Computational Performance

Our framework achieves superior accuracy while maintaining computational efficiency suitable for operational deployment. Training on 11 years of data takes 22 hours on a single NVIDIA A100 GPU. This is much faster than the 71 hours required by FLake NWP's data assimilation and the 100 hours required by MetNet-3's larger architecture. Thanks to its modular design, the framework can be updated incrementally as new data becomes available. Incorporating an additional month of observations, for example, requires only two hours.

The inference time meets operational requirements, executing a complete 72-hour forecast in 15 seconds on standard hardware. The adaptive targeting mechanism significantly contributes to this efficiency by processing high-resolution predictions only where needed. Memory requirements peak at 24 GB during inference, enabling deployment on current-generation operational systems without specialized hardware.

5.9 Discussion and Limitations

Our evaluation reveals that the combination of data synthesis, temporal pattern recognition, physical constraints, and adaptive resolution successfully addresses the key challenges in predicting lake-effect snow. The framework's superior performance does not stem from any single innovation, but rather from the careful integration of complementary approaches that address different aspects of the prediction problem.

There are several limitations that remain for future work. Complex terrain interactions, particularly in the Michigan Upper Peninsula, sometimes produce precipitation patterns that our model has difficulty capturing. The fixed 11-year training period may not fully represent climate variability, suggesting the benefits of continual learning approaches. Transitions between lake-effect and synoptic snow remain challenging because these events involve interactions across scales that are beyond the scope of our current modeling framework.

Despite these limitations, our hybrid approach is a significant advancement in lake-effect snow prediction. It provides accurate, physically consistent forecasts at the required spatial and temporal scales for effective hazard mitigation.

6 Conclusion

This work demonstrates that solving fundamental data limitations can unlock the full potential of physics-informed machine learning for environmental prediction. By addressing the temporal discontinuity in satellite observations—a challenge that has constrained lake-effect snow forecasting for decades—we enable improved prediction models that combine physical understanding with data-driven learning.

Our two-stage framework represents a novel approach to handling observational gaps in meteorology. Rather than developing increasingly sophisticated models to work around missing data, we first reconstruct complete observational sequences through cross-spectral synthesis. The PatchGAN approach achieves remarkable fidelity in generating nighttime visible and near-infrared imagery from continuous infrared observations, maintaining both visual quality (SSIM 0.82) and meteorological consistency. This synthesis alone improves downstream prediction accuracy by 59%, validating our hypothesis that temporal completeness is essential for capturing atmospheric evolution.

Based on full observations, our physics-informed architecture provides surprising insights into lake-effect predictability. The dramatic improvement in harsh event detection, from 27.1% at 24 hours to 77.6% at 72 hours, challenges the notion that forecasts degrade over time. Our findings suggest that severe lake-effect events are preceded by large-scale atmospheric patterns that become increasingly apparent over multiday timescales, but only when models have access to continuous observations that capture these evolving signatures. Integrating conservation laws and thermodynamic constraints through the PINN module ensures that these extended predictions remain physically plausible, which addresses a key limitation of purely statistical approaches.

From an operational perspective, our framework provides weather services and emergency management with immediate benefits. The adaptive spatial targeting reduces computational requirements by 65-80% while maintaining a 500-meter resolution in critical zones. This makes deployment feasible on current operational infrastructure. With a mean spatial error of 8.6 km, predictions accurately identify affected communities, which is crucial for public safety when neighboring towns can experience drastically different conditions. The extension of reliable forecasts from 18 to 72 hours gives emergency managers more time to prepare for severe events.

Several limitations warrant acknowledgment and future investigation. First, our framework exhibits reduced performance when transitioning between lake-effect and synoptic snow, as scale interactions surpass the current modeling capabilities. The fixed training period may not fully capture climate variability, suggesting the benefits of continual learning approaches. Complex terrain effects, particularly in the Michigan Upper Peninsula, occasionally produce precipitation patterns that our model struggles to predict accurately. Additionally, while our synthesis approach works well for the considered spectral bands, extending it to other observational modalities requires further research.

Generalizability Across the Great Lakes Region: Our evaluation focuses exclusively on Lake Michigan, which limits its claims about generalizability to other Great Lakes or similar water bodies worldwide. Lake-effect dynamics exhibit significant variation across the Great Lakes system due to differences in:

- **Lake geometry:** Lake Michigan's north-south orientation creates different fetch patterns compared to the east-west elongation of Lake Erie or the massive size of Lake Superior
- **Surrounding topography:** The relatively flat terrain around Lake Michigan differs markedly from the complex topography around Lake Ontario or the Appalachian influences on Lake Erie
- **Urban heat islands:** The Chicago metropolitan area significantly affects local atmospheric conditions in ways that may not apply to other lake regions
- **Climatological patterns:** Each lake experiences different seasonal ice coverage, temperature regimes, and prevailing wind patterns

While our physics-informed constraints should transfer across lakes (fundamental atmospheric laws remain constant), the learned patterns in both the PatchGAN synthesis and ConvLSTM components may be lake-specific. The adaptive targeting thresholds (α weights, decay lengths, resolution breakpoints) were optimized for Lake Michigan's characteristics and would likely require recalibration for other lakes.

Initial analysis suggests that Lakes Huron and Superior, with similar size scales and surrounding terrain, might require minimal adaptation. However, Lakes Erie and Ontario, with their distinct morphologies and more complex surrounding topography, could necessitate substantial model retraining. Transfer learning approaches could potentially reduce the data requirements for adapting to new lakes, but this remains untested.

Regional Climate Considerations: Our 11-year training period (2006-2017) may not fully capture the range of climate variability affecting lake-effect patterns. Longer-term climate shifts, such as changing ice coverage patterns due to warming temperatures or evolving storm tracks, could impact model performance. The framework would benefit from continual learning capabilities that adapt to changing climate conditions while preserving learned physical relationships.

Looking ahead, this work suggests several promising research directions. The success of cross-spectral synthesis suggests that similar approaches could address observational gaps in other remote sensing applications, ranging from wildfire monitoring to agricultural assessment. The framework's architecture can be naturally extended to other Great Lakes or similar bodies of water, though transfer learning strategies still need to be developed. Integrating the framework with ensemble prediction systems could quantify uncertainty in the synthesis and prediction stages. Most intriguingly, the counterintuitive improvement in long-range harsh event prediction merits deeper investigation into the atmospheric dynamics enabling this extended predictability.

Beyond its technical contributions, this work highlights the importance of challenging fundamental assumptions in environmental prediction. The long-standing acceptance of nighttime observational gaps as an unavoidable limitation has led to increasingly complex workarounds. Addressing this root cause directly improves lake-effect snow prediction and establishes a template for solving other challenging forecasting problems where sparse observations, fine-scale dynamics, and physical constraints intersect. As climate change intensifies extreme weather events, a holistic approach combining data synthesis, physics-informed learning, and adaptive computation will be critical to protecting vulnerable communities.

References

[1] F. Alyahyai. Tailored modeling techniques for lake-effect snow events. *Advances in Meteorological Science*, 30(4):3192–3204, 2010.

[2] M. Berenguer, M. Surcel, I. Zawadzki, M. Xue, and F. Kong. The diurnal cycle of precipitation from continental radar mosaics and numerical weather prediction models. Part II: Intercomparison among numerical models and with nowcasting. *Monthly Weather Review*, 140(8):2689–2705, 2012.

[3] L. Besombes and coauthors. Producing realistic climate data with generative adversarial networks. *Natural Hazards and Earth System Sciences*, 28(3):347–359, 2021.

[4] G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, Englewood Cliffs, NJ, 1970.

[5] T. Can. Integration of multiple meteorological data sources for improved forecasting. *International Journal of Meteorological Research*, 10(2):101–110, 2017.

[6] X. Chen and coauthors. Change-point analysis as a tool to detect abrupt climate variations. *International Journal of Climatology*, 36(4):200–210, 2016.

[7] Michael Dixon and Gerry Wiener. Titan: Thunderstorm identification, tracking, analysis, and nowcasting—a radar-based methodology. *Journal of Atmospheric and Oceanic Technology*, 10(6):785–797, 1993.

[8] Shuran Gao, Ming Cheng, Kangkang Zhao, Xiyang Zhang, Jian Han, Jing Liu, and Xiang Bai. Res2net: A new multi-scale backbone architecture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 630–638, 2019.

[9] A. J. Geer, F. Baordo, N. Bormann, P. Chambon, S. J. English, M. Kazumori, H. Lawrence, P. Lean, K. Lonitz, and C. Lupu. The growing impact of satellite observations sensitive to humidity, cloud and precipitation. *Quarterly Journal of the Royal Meteorological Society*, 143(709):3189–3206, 2017.

[10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 2672–2680, 2014.

[11] Great Lakes Environmental Research Laboratory (GLERL). Coastwatch program – great lakes environmental research laboratory. `https://coastwatch.glerl.noaa.gov/`, 2025. Daily satellite-derived measurements at 1.8 km resolution; Accessed: 2025-02-28.

[12] Great Lakes Environmental Research Laboratory (GLERL). Real-time monitoring buoy network. `https://www.glerl.noaa.gov/`, 2025. Temperature profiles and water pressure measurements via instrumented buoys; Accessed: 2025-02-28.

[13] M. Gyamerah and coauthors. Regime-switching temperature dynamics model for weather derivatives. *arXiv preprint arXiv:1808.04710*, 2018.

[14] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[15] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017.

[16] L. Ji and coauthors. Time series prediction method for meteorological data based on the arima-lstm model. *Academic Journal of Science and Technology*, 10(1):100–110, 2024.

[17] M. et al. Johannsen. Evaluation of nowcasting techniques for short-term weather prediction. *Journal of Forecasting*, 39(4):350–366, 2020.

[18] E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, Cambridge, UK, 2003.

[19] K. Lee and R. Patel. Titan: A novel storm-tracking algorithm for radar data. In *Proceedings of the IEEE International Conference on Geoscience and Remote Sensing*, pages 1425–1429, 2020.

[20] P. Li and coauthors. Precipitation nowcasting using diffusion transformer with causal attention. *arXiv preprint arXiv:2410.13314*, 2023.

[21] E. N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, 1963.

[22] John D. McMillen and W. James Steenburgh. Capabilities and limitations of convection-permitting wrf simulations of lake-effect systems over the great salt lake. *Weather and Forecasting*, 30(6):1711–1731, 2015.

[23] R. Miller. The influence of vegetation on lake-effect snow distribution. *Journal of Hydrometeorology*, 5(2):210–221, 2004.

[24] National Oceanic and Atmospheric Administration (NOAA). Goes satellite program. `https://www.nesdis.noaa.gov/GOES`, 2025. Accessed: 2025-02-28.

[25] National Weather Service. Meteorological datasets. `https://www.weather.gov/`, 2025. Extensive meteorological measurements including temperature, wind speed, wind chill, and heat index; Accessed: 2025-02-28.

[26] D. Niziol. State and evolution of lake-effect snow: Impacts and challenges. *Journal of Atmospheric Research*, 82(3):123–135, 2008.

[27] R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer, New York, NY, 4th edition, 2017.

[28] L. et al. Song. Deep learning approaches for localized weather prediction. *Remote Sensing of Environment*, 240:111—-121, 2020.

[29] H. Vieus and F. Dupont. Hydrological considerations in lake-effect snow prediction. *Hydrology and Earth System Sciences*, 8(3):549–560, 2004.

[30] K. W. Wong et al. Using wavelets for time series forecasting: Does it pay off? *Economic Modelling*, 20(2):123–130, 2003.

[31] Lufei Zhao, Tonglin Luo, Xuchu Jiang, and Biao Zhang. Prediction of soil moisture using bigru-lstm model with stl decomposition in qinghai–tibet plateau. *PeerJ*, 11:e15851, 2023.