Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

AI-Assisted Link Quality Prediction for Weather-Aware Outdoor Terahertz Communications: Forecasting, Uncertainty, and Explainability Analysis

Israt Jahan

isratjahan@cuet.ac.bd

Chittagong University of Engineering and Technology (CUET)

Md. Jahedul Islam

Chittagong University of Engineering and Technology (CUET)

Research Article

Keywords: Terahertz communications, rain attenuation, link quality prediction, machine learning, uncertainty quantification, explainable AI

Posted Date: September 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11128433/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

AI-Assisted Link Quality Prediction for Weather-Aware Outdoor Terahertz Communications: Forecasting, Uncertainty, and Explainability Analysis

Israt Jahan [1*] and Md. Jahedul Islam [1*]

[1*]Department of Electronics and Telecommunication Engineering, Chittagong University of Engineering and Technology (CUET), Chattogram, Bangladesh.

*Corresponding author(s). E-mail(s): isratjahan@cuet.ac.bd; jahed@cuet.ac.bd;

Abstract

Outdoor terahertz (THz) communication links suffer severe, time-varying degradation from rain and atmospheric attenuation, yet most prior work including our earlier study evaluates this impairment at fixed, discrete rain rates rather than forecasting it as a dynamic process. This paper proposes a real-weather-driven THz channel and Link Quality Index (LQI) forecasting framework: a composite, application-relevant metric combining channel, reliability, and hardware margins, computed by a physics-based simulator (ITU-R P.838-3 rain attenuation, ITU-R P.676 gaseous attenuation, and a modified Friis link budget) driven by six years of hourly meteorological observations (51,888 records) from Sylhet , Bangladesh. We benchmark seven forecasting models and these are persistence, linear regression, random forest, XGBoost, GRU, LSTM, and a Transformer across four forecast horizons (1, 3, 6, and 12 hours) under a leak-free chronological 70/15/15 split, with hyperparameter sensitivity, multi-seed robustness, Diebold–Mariano significance testing, Monte Carlo Dropout uncertainty calibration, SHAP explainability, a feature-group ablation, and a degradation-conditioned stratified evaluation. Results show forecast skill of 21.5%–31.4% over persistence, well-calibrated but progressively overconfident uncertainty at longer horizons, and substantially higher error during rain and low-LQI conditions. To further validate the framework, we apply the identical pipeline to a second, climatically contrasting site (Rajshahi, Bangladesh) and find that the central

conclusions, including that aggregate metrics hide severe forecast failures under degraded conditions, hold consistently across both sites.

Keywords: Terahertz communications, rain attenuation, link quality prediction, machine learning, uncertainty quantification, explainable AI

1 Introduction

Terahertz (THz) communication (0.1–10 THz) is a leading candidate physical-layer technology for sixth-generation (6G) wireless networks [1, 19, 28], offering enormous bandwidth to satisfy the terabit-per-second data-rate, sub-millisecond-latency, and ubiquitous-connectivity targets envisioned for 6G [12, 54, 60, 63, 66, 67]. This enormous bandwidth, however, comes at the cost of severe free-space path loss and atmospheric attenuation, of which rain is the most critical impairment: it can drive the bit error rate (BER) to unacceptable levels within a few metres of propagation and is a key open challenge identified across recent THz channel surveys [21, 33, 55]. Artificial intelligence is increasingly positioned as a native design principle for 6G physical layers rather than an external add-on [37, 47, 57], motivating the forecasting-first approach taken here.

Our earlier conference study characterized this impairment using the ITU-R P.838-3 rain attenuation model [27] combined with a modified Friis equation, evaluating link performance at fixed, discrete rain-rate snapshots [32]. While informative, that characterization – and the great majority of related THz rain-attenuation literature is static and reactive: it reports what happens if a given rain rate occurs, rather than forecasting weather-driven degradation ahead of time. A parallel body of work has begun exploring machine-learning-based forecasting of weather-driven link degradation in adjacent bands, including deep-learning rain-fade prediction for satellite links [16] and microwave-link attenuation forecasting [31], camera-imagery-driven link-quality prediction for millimeter-wave/THz handover [30], LSTM-based wireless link-quality estimation [34], and AI-driven predictive gateway-diversity switching for rain-affected satellite systems [10]; however, these efforts either operate at lower frequencies, rely on non-meteorological sensing modalities, or do not couple a physically grounded, standards-based THz channel model with a rigorous, leak-free forecasting benchmark, uncertainty quantification, and explainability analysis in a single pipeline. Composite link-quality metrics have likewise been explored primarily for lower-frequency wireless sensor networks [42] rather than as a forecastable, physically anchored quantity for outdoor THz links, and explainable AI has seen growing adoption elsewhere in wireless system design [20, 35] but not yet for weather-driven THz link forecasting specifically. Table 1 summarizes this positioning against the most closely related prior studies.

This paper moves from a static, reactive characterization toward a predictive one: can machine learning forecast weather-driven THz link degradation ahead of time, using real meteorological time series rather than fixed scenarios? We introduce the Link Quality Index (LQI), a composite, normalized, forecastable metric distinct from the field's prevailing focus on predicting raw BER or signal-to-noise

Table 1 Positioning against the most closely related prior work. Columns: Dyn. Input = dynamic, time-varying sensing input (weather, imagery, or signal history); Comp. = composite QoS metric; ML Fc. = ML forecasting (vs. static); M-H = multi-horizon (1–12 h); UQ = uncertainty quantified; XAI = explainability (SHAP/LIME).

Study	Band/Domain	Dyn. Input	Comp.	ML Fc.	M-H	UQ	XAI
Jahan and Islam [32] (prior work)	THz, 300 GHz	No (fixed rain)	No (raw BER/SNR)	No	No	No	No
Hossain and Islam [26]	Satellite, EHF	No (fixed rain)	No	No	No	No	No
Ferdowsi and Whitefield [16]	Satellite, Ka-band	Yes	No	Yes	No	No	No
Jacoby et al. [31]	Microwave link	Yes	No (atten. only)	Yes	No	No	No
Itahara et al. [30]	mmWave/THz handover	Yes (camera)	No	Yes	No	No	No
Kanto and Watabe [34]	General wireless	Yes (RSSI)	Partial (LQE)	Yes	No	No	No
This work	THz, 300 GHz	Yes (6-yr, hourly)	Yes (Ch.+Rel.+HW)	Yes (7models)	Yes	Yes (MCD)	Yes (SHAP)

ratio (SNR) in isolation, and benchmark a full forecasting pipeline – from a naive persistence baseline through gradient-boosted trees and recurrent/attention-based deep sequence models with uncertainty quantification via Monte Carlo (MC) Dropout [18] and explainability via SHAP [43], together with a set of robustness studies (hyperparameter tuning, multi-seed variance, a Diebold–Mariano point-wise significance test, an attention-based Transformer baseline [62], a feature-group ablation, and a degradation-conditioned stratified evaluation).

1.1 Contributions

The contributions of this paper are as follows.

- A real-weather-driven THz channel and LQI dataset (six years, hourly resolution, Sylhet, Bangladesh) – a reusable artefact distinct from the fixed-scenario characterizations common in prior work [32].
- The Link Quality Index (LQI): a composite, application-relevant, physically anchored metric combining channel, reliability, and hardware margins, with a transparently documented and corrected collinearity issue.
- A six-model forecasting benchmark (persistence, linear regression, random forest, XGBoost [11], GRU [13], LSTM [24]) across four horizons under a strictly leak-free chronological split, supported by hyperparameter tuning, multi-seed robustness, and a Diebold–Mariano point-wise significance test.
- A lightweight attention-based Transformer baseline [62], trained under an identical protocol to the recurrent models, to test whether attention-based sequence modeling offers additional forecast skill at this problem scale.
- Uncertainty-aware (MC Dropout [18]) and explainable (SHAP [43]) prediction, including an honest report of calibration degradation with forecast horizon.
- A feature-group ablation isolating the marginal value of weather history versus LQI autocorrelation, and a degradation-conditioned stratified evaluation showing that

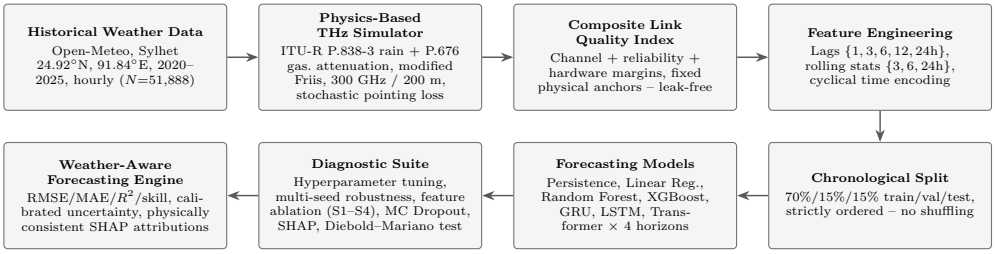

Fig. 1 Weather-aware LQI forecasting pipeline. The flow reads left to right along the top row and continues right to left along the bottom row.

aggregate metrics substantially understate forecast error during the rain/degraded conditions that motivate this research.

- Cross-site validation (Section 5.2) applying the identical pipeline, unchanged, to a second, climatically contrasting site (Rajshahi), confirming that the central findings generalize beyond a single location.

We emphasize at the outset that the contribution here is methodological rigor and system-level integration – a physics-grounded, leak-free, uncertainty-aware, and explainable forecasting pipeline built on a genuine multi-year weather record rather than a new learning algorithm; all forecasting models used are established architectures. We view this as the right contribution for a physical-layer/systems venue: Section 5 explains why simpler, off-the-shelf models are the correct engineering choice at this problem scale, and why the sequence-model and uncertainty-quantification machinery is nonetheless retained.

The remainder of this paper is organized as follows. Section 2 describes the THz channel simulator based on physics and the definition of the LQI. Section 3 details the forecasting methodology, feature engineering, and the evaluation protocol. Section 4 reports the forecast results, robustness studies, uncertainty and explainability analysis, ablation, and degradation-specific evaluation. Section 5 discusses the key findings and limitations, and Section 6 concludes the paper.

2 System Model: THz Channel Simulator and the Link Quality Index

Rather than driving the channel model with arbitrarily chosen fixed rain rates as in [32], the simulator used in this work is driven by real, time-stamped meteorological variables (namely rain rate, humidity, temperature, and pressure) sourced from an Open-Meteo reanalysis series [68] for Sylhet, Bangladesh (24.92°N, 91.84°E), spanning 51,888 hourly records (∼6 years, 2020–2025).These variables are propagated through a physics-based simulator combining ITU-R P.838-3 rain attenuation [27], ITU-R P.676 gaseous attenuation [29], and a modified Friis link budget to produce a time series of channel conditions and the derived LQI. The overall pipeline is summarized in Fig. 1.

Table 2 Link budget and system parameters.

Parameter	Value
Operating frequency	300 GHz
Link distance	200 m
TX power	13 dBm
TX / RX antenna gain	35 / 35 dBi
Noise figure	8 dB
Bandwidth	10 GHz
Pointing-loss impairment	$\mathcal{N}(1.5, 1.0)$ dB, clipped $[0, 4]$
Polarization / elevation	45° (circular) / 0° (horizontal)

2.1 Link Budget and System Parameters

Table 2 summarizes the link-budget and system parameters used by the simulator.

Each parameter reflects a plausible, currently reportable outdoor THz link rather than an idealized or arbitrarily chosen one. The 300 GHz carrier sits in one of the low-absorption windows actively targeted for 6G THz research and matches our own prior work [32], allowing direct comparison. A 200 m span represents a realistic short-range fixed link, such as a last-mile or inter-building backhaul hop. It is long enough for weather to matter, yet short enough for THz to remain viable.In addition outdoor 300 GHz links of this length and longer are already demonstrated in practice, including a 200 m, 2×2 MIMO 300 GHz link at over 200 Gb/s [9]. 13 dBm TX power is representative of what current THz front-end amplifiers can deliver; THz hardware remains power-limited relative to lower-frequency RF, so this figure was not chosen to make the link look better than today's components allow. The 35 dBi TX/RX antenna gains reflect the high-gain directional antennas, such as phased-array or lens antennas, that are essentially mandatory at this band to overcome the enormous free-space path loss. In fact, this value is conservative compared to demonstrated 300 GHz Cassegrain antennas, which reach as high as 40 dBi[39]. The 8 dB noise figure, by contrast, sits at the optimistic edge of current state-of-the-art hardware. For instance, the best reported standalone low-noise amplifiers near 300 GHz achieve 10.5 dB [65], so the chosen value reflects a near-term achievable target for an integrated receiver chain rather than today's median demonstrated performance. 10 GHz of bandwidth is a deliberately conservative slice of the many-GHz windows THz makes available, chosen to illustrate the bandwidth advantage without assuming an unrealistically wide channel. The stochastic pointing-loss model captures unavoidable beam-misalignment jitter in a narrow, highly directional beam, independent of weather. Circular polarization at 45° reduces sensitivity to transmitter/receiver rotational misalignment, and 0° elevation reflects the horizontal, ground-level, fixed point-to-point geometry assumed throughout.

Rain-induced attenuation is computed from the ITU-R P.838-3 specific-attenuation model [27], in which the specific attenuation γ_R (dB/km) follows a power law in the rain rate R (mm/h), with frequency- and polarization-dependent coefficients k and α:

$$\gamma_R = k \cdot R^{\alpha}. \tag{1}$$

The total rain attenuation over the link is obtained by scaling the specific attenuation by the link distance d (km):

$$A_{\text{rain}} = \gamma_R \cdot d. \tag{2}$$

For this link geometry at 300 GHz, the fitted ITU-R P.838-3 coefficients are $k = 1.629$ and $\alpha = 0.628$, giving a rain-attenuation range of 0.000–3.318 dB across the six-year sample.

Gaseous (water-vapour and oxygen) attenuation follows ITU-R P.676 [29], driven by temperature, relative humidity, and pressure. Water-vapour density ρ_v (g/m^3) is obtained from the Magnus formula for saturation vapour pressure e_s (hPa) and actual vapour pressure e:

$$e_s = 6.1094 \exp\left(\frac{17.625\,T}{T + 243.04}\right), \tag{3}$$

$$\rho_v = 216.7 \cdot \frac{RH}{100} \cdot \frac{e_s}{T + 273.15}, \tag{4}$$

where T is air temperature (°C) and RH is relative humidity (%). The resulting gaseous attenuation A_{gas} is obtained from the ITU-R P.676 terrestrial-path model using ρ_v, pressure, and link distance, giving a range of 0.711–4.522 dB across the sample (total attenuation $A_{\text{rain}} + A_{\text{gas}} = 0.711$–$6.463$ dB).

Free-space path loss follows the standard Friis relation for distance d (m) and frequency f (Hz), with c the speed of light:

$$\text{FSPL} = 20\log_{10}(d) + 20\log_{10}(f) + 20\log_{10}\left(\frac{4\pi}{c}\right). \tag{5}$$

Received power is given by a modified Friis link-budget equation in which the free-space term is augmented by the rain, gaseous, and pointing-loss impairments:

$$P_{\text{RX}} = P_{\text{TX}} + G_{\text{TX}} + G_{\text{RX}} - \text{FSPL} - A_{\text{rain}} - A_{\text{gas}} - L_p, \tag{6}$$

with L_p the stochastic pointing/misalignment loss (Table 2).

Thermal noise power N (dBm) and the resulting SNR follow the standard relations, with k_B Boltzmann's constant, T_K the system temperature (290 K), B the bandwidth, and NF the receiver noise figure:

$$N = 10\log_{10}(k_B T_K B \cdot 1000) + NF, \tag{7}$$

$$\text{SNR} = P_{\text{RX}} - N. \tag{8}$$

Bit error rate is evaluated per modulation scheme from the resulting SNR (per-symbol, E_s/N_0) using the standard additive white Gaussian noise (AWGN) expressions for on-off keying (OOK), quadrature phase-shift keying (QPSK), and M-ary quadrature amplitude modulation (QAM, $\log_2 M$ bits/symbol):

$$\text{BER}_{\text{OOK}} = \tfrac{1}{2}\,\text{erfc}\sqrt{\text{SNR}/4}, \tag{9}$$

$$\mathrm{BER}_{\mathrm{QPSK}} = \tfrac{1}{2}\,\mathrm{erfc}\sqrt{\mathrm{SNR}/2}, \tag{10}$$

$$\mathrm{BER}_{M\text{-QAM}} = \frac{4}{\log_2 M}\left(1 - \frac{1}{\sqrt{M}}\right) \times \tfrac{1}{2}\,\mathrm{erfc}\sqrt{\frac{3\,\mathrm{SNR}}{2(M-1)}}. \tag{11}$$

The transmitter adaptively selects the highest-order scheme whose predicted BER still satisfies a target BER of 10^{-6}, falling back to OOK when no scheme meets the target. Across this link's observed SNR range (11.3–20.4 dB over both six-year records), QPSK's BER clears this target almost everywhere, giving it the practical role of the link's workhorse scheme; 16-QAM's own threshold (≈20.4 dB) is missed by only ≈0.3 dB at the single best moment in either record, a near-miss rather than a structural impossibility, so it is retained as a candidate scheme even though it is never actually selected in the resulting dataset; 64-QAM's threshold (≈26.6 dB) is roughly 6 dB beyond anything observed and is excluded from the candidate list entirely on that basis.

2.2 The Link Quality Index (LQI)

Most THz rain-attenuation studies, including [32], report BER or SNR in isolation. In a deployed system, however, engineers care whether the link can reliably support the running application, so the LQI is defined as a weighted, normalized composite of three physically anchored margins: a channel margin (received power relative to a fixed outage point and a fixed clear-sky reference), a reliability margin (QPSK BER relative to fixed log-scale anchors), and a hardware margin (pointing/misalignment loss relative to its defined maximum, independent of weather).

Each raw quantity x is mapped to a component margin via fixed min–max normalization against its worst and best physical anchors, clipped to $[0, 1]$:

$$\mathrm{norm}(x) = \mathrm{clip}\left(\frac{x - x_{\mathrm{worst}}}{x_{\mathrm{best}} - x_{\mathrm{worst}}},\, 0,\, 1\right), \tag{12}$$

with the reliability margin computed on the log-BER scale so that the many-orders-of-magnitude spread in BER is compressed to a usable numeric range:

$$M_{\mathrm{reliability}} = \mathrm{norm}(-\log_{10}(\mathrm{BER}_{\mathrm{QPSK}}))\,. \tag{13}$$

The three normalized margins are combined into the composite LQI via a fixed weighted sum:

$$\mathrm{LQI} = w_c M_{\mathrm{channel}} + w_r M_{\mathrm{reliability}} + w_h M_{\mathrm{hardware}}, \tag{14}$$

with weights $w_c = 0.45$, $w_r = 0.40$, $w_h = 0.15$ (Table 3). All anchors are fixed physical reference points rather than sample percentiles, so LQI values computed on any subset of years are bit-for-bit identical to the same rows extracted from the full six-year run.

Table 3 LQI component weights and anchors.

Component	Weight	Worst Anchor	Best Anchor
Channel margin	0.45	-53.165 dBm	-46.544 dBm
Reliability margin	0.40	BER $= 10^{-3}$	BER $= 10^{-15}$
Hardware margin	0.15	4 dB	0 dB

Table 4 LQI weight sensitivity: alternative weight schemes.

Scheme ($w_c/w_r/w_h$)	corr(rain)	corr($\mathrm{LQI_o}$)	Spearm. ρ
Original (.45/.40/.15)	-0.490	1.000	1.000
Equal (.33/.33/.34)	-0.370	0.900	0.901
Channel-heavy (.60/.30/.10)	-0.489	0.990	0.989
Reliab.-heavy (.30/.60/.10)	-0.530	0.990	0.996
Hardware-heavy (.35/.35/.30)	-0.398	0.931	0.933

A companion column, $\mathrm{LQI_{weather\text{-}only}}$ (identical formula with the hardware term dropped and renormalized), is retained in the released dataset. It represents the ceiling a purely weather-driven forecaster could ever reach, since approximately 11% of full-LQI variance originates from pointing-loss noise that no weather-only model can predict; this is revisited as a limitation in Section 5.

2.3 Weight Sensitivity

The weights $w_c = 0.45$, $w_r = 0.40$, $w_h = 0.15$ are physically motivated rather than fit to the data, but a composite index's specific weighting can still appear arbitrary without a sensitivity check. Table 4 recomputes LQI under four alternative weight schemes directly from the simulator's stored channel, reliability, and hardware margins, and reports each variant's correlation with rain rate and total attenuation, its Pearson correlation with the original LQI, and the Spearman rank correlation of its hourly ordering against the original.

When only the channel and reliability weights are adjusted, and the hardware weight is kept at or below its original 0.15, the hourly LQI ordering barely changes (Spearman $\rho \geq 0.989$). Both the Channel-heavy and Reliability-heavy schemes stay nearly identical to the original in rank and in their correlation with rain rate. The picture changes, however, once a scheme inflates the hardware weight instead: the Equal and Hardware-heavy schemes both show weaker rain-rate correlation and weaker rank agreement. This happens because the hardware margin is, by construction, weather-independent pointing-loss noise, so giving it more weight simply dilutes the weather signal the index is meant to carry. This result is the quantitative basis for keeping w_h low: the exact value 0.15 is not what matters, but keeping it small relative to w_c and w_r is. All forecasting results in Section 4 are computed using the original, physically motivated weighting throughout. In short, $w_c = 0.45$, $w_r = 0.40$, and $w_h = 0.15$ reflect a deliberate design choice rather than a fitted one. Channel and reliability margins

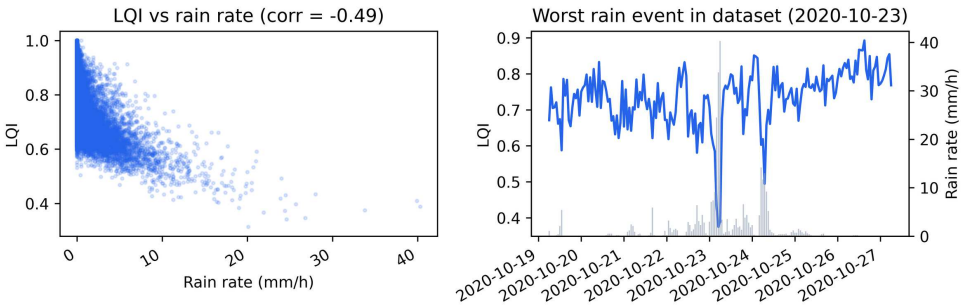

Fig. 2 Simulator validation: LQI vs. rain rate across the full six-year sample (correlation = −0.49) and the most severe rain event on record, showing LQI tracking rain intensity in time.

are weighted comparably and dominate the index because they most directly determine whether the link can sustain a target application, while the hardware margin is given a smaller weight since pointing/misalignment loss is weather-independent by construction and should influence, but not dominate, a weather-forecasting-oriented metric. Table 4 establishes the quantitative robustness of this choice, but it is worth being precise about what that robustness check does and does not show: no forecasting model was retrained under the alternative weightings, so no claim is made about how a model would perform under them. The check is limited to the index's own internal consistency, meaning how stable its correlations and rankings are, which is exactly what Table 4 reports.

2.4 LQI Validation: Why the Hardware Margin Matters

Including the hardware margin in the final weighted sum, rather than letting the channel and reliability margins alone determine LQI, keeps the correlation between LQI and total attenuation at −0.925; without it, LQI would be little more than a rescaling of total attenuation ($\rho \approx -0.999$), leaving the forecasting task in Section 4 correspondingly less interesting. The reliability margin's "excellent" BER anchor is likewise set at 10^{-15} rather than a looser threshold such as 10^{-9}, which would leave it saturated near 1.0 for over 99% of the sample and contributing little useful information.

Fig. 2 confirms that LQI tracks rain intensity in time, including during the most severe rain event on record in the sample, and Fig. 3 confirms monotonic, physically sensible relationships between LQI and each intermediate physical quantity (rain rate, received power, SNR, BER, and total attenuation), validating that the simulator behaves as the underlying physics dictates.

2.5 Resulting Dataset Summary

Table 5 summarizes the resulting six-year dataset. The LQI spans the full $[0.31, 1.00]$ range with mean 0.816 (std. 0.106), is correlated with rain rate at −0.49 and with total attenuation at −0.925, and induces an adaptive-modulation distribution in which QPSK covers the overwhelming majority of hours (51,406), with OOK reserved for

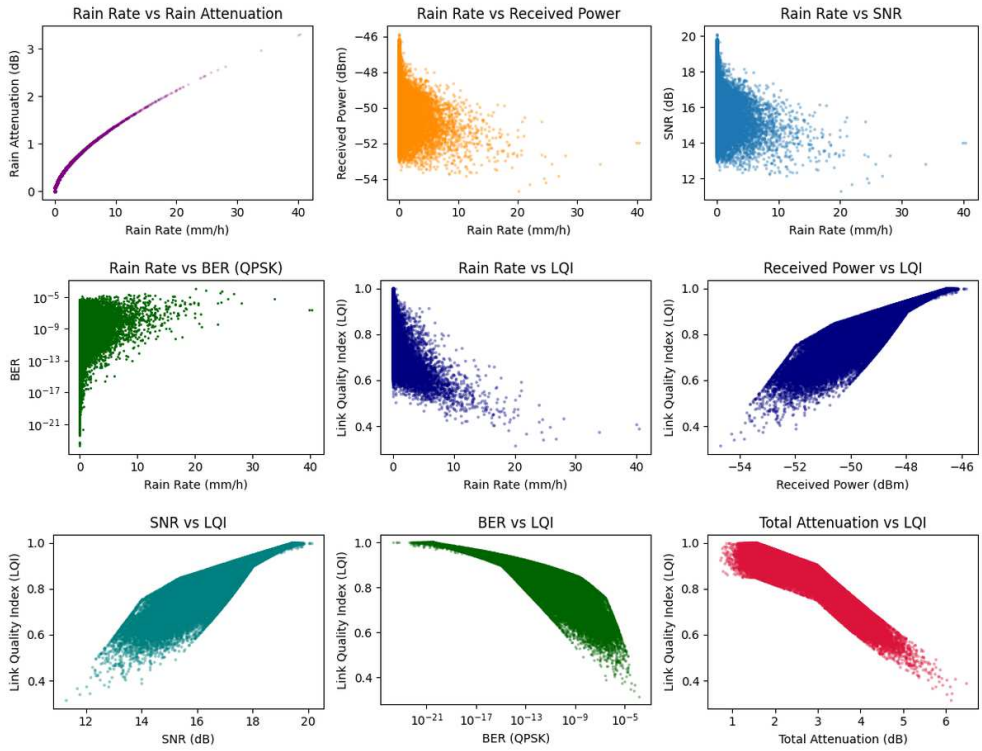

Fig. 3 Physical-consistency diagnostic suite: rain rate vs. attenuation/received power/SNR/BER, and LQI vs. each intermediate physical quantity. Monotonic, physically sensible relationships throughout confirm the simulator behaves as the underlying physics dictates.

Table 5 Resulting dataset summary.

Statistic	Value
Rows (hourly)	51,888 (∼6 years)
LQI range	0.3138–1.0000
LQI mean (std.)	0.8161 (0.1062)
corr(LQI, rain rate)	−0.490
corr(LQI, total attenuation)	−0.925
Adaptive modulation (QPSK/OOK)	51,406 / 482

the most degraded conditions (482 hours); 16-QAM is never selected within this link's observed SNR range, for the physical reasons discussed above.

3 Forecasting Methodology

3.1 Feature Engineering

From the LQI dataset, three feature groups are engineered for every forecast horizon $h \in \{1, 3, 6, 12\}$ hours. The first is *instantaneous weather and cyclical time* (8 features): rain rate, temperature, humidity, pressure, and sine/cosine encodings of hour-of-day and day-of-year. The second is *weather history* (14 features): rain and humidity lags at $\{1, 3, 6, 12, 24\}$ h, plus rolling mean/max statistics over $\{3, 6, 24\}$ h windows. The third is *LQI history* (5 features): LQI lags at $\{1, 3, 6, 12, 24\}$ h. Together these give a total of 27 features for the classical/XGBoost models [11, 17]. The GRU [13], LSTM [24], and Transformer [62] sequence models instead consume a 24-hour lookback window of five raw channels (rain rate, temperature, humidity, pressure, LQI) per timestep, following standard lag/rolling-window feature-engineering practice for weather-driven time-series forecasting [4, 14].

3.2 Leak-Free Chronological Split

All experiments use a strict 70%/15%/15% chronological split (train/validation/test) with no shuffling which is a standard practice for time-series forecasting, since random shuffling would let a model see future information during training and inflate its apparent skill [3]. The test set is touched exactly once per experiment, at final evaluation; all model and hyperparameter selection uses validation-set performance only. The `StandardScaler` [49] used for the GRU/LSTM/Transformer's continuous channels is fit on the training split only, never on validation or test data.

3.3 Models and Evaluation Protocol

Six baseline models are benchmarked – persistence (a naive "value stays the same" forecast), linear regression, random forest [7, 8], XGBoost [11], GRU [13], and LSTM [24] and each was trained independently per horizon, together with the attention-based Transformer baseline introduced in Section 4.2.1 below. GRU and LSTM are chosen as the recurrent baselines following their established effectiveness for time-series forecasting broadly [50] and for wireless-channel forecasting specifically [59]. XGBoost uses early stopping on the validation set (500 estimators, depth 6, learning rate 0.05, subsample/colsample 0.8); default hyperparameters are validated against a grid search in Section 4.3. GRU and LSTM each use a single 32-unit recurrent layer, the Adam optimizer [36] (learning rate 10^{-3}), batch size 512, and up to 15 epochs with early stopping (patience 3) on validation loss; implementation uses PyTorch [48].

For N test points with true value y and forecast $\hat{y}$, three standard regression metrics are reported, together with forecast skill relative to the persistence baseline:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_i (y_i - \hat{y}_i)^2}, \tag{15}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_i |y_i - \hat{y}_i|, \tag{16}$$

Table 6 Full model comparison (seed= 42), all four horizons.

h (h)	Model	RMSE	MAE	Skill
1	Persistence	0.0620	0.0473	0.0%
1	Linear Regression	0.0514	0.0393	17.0%
1	Random Forest	0.0502	0.0385	19.0%
1	XGBoost	0.0487	0.0371	21.5%
1	GRU	0.0491	0.0373	20.8%
1	LSTM	0.0493	0.0377	20.5%
3	Persistence	0.0712	0.0548	0.0%
3	Linear Regression	0.0551	0.0421	22.6%
3	Random Forest	0.0546	0.0416	23.3%
3	XGBoost	0.0538	0.0409	24.4%
3	GRU	0.0538	0.0410	24.3%
3	LSTM	0.0546	0.0417	23.3%
6	Persistence	0.0769	0.0589	0.0%
6	Linear Regression	0.0567	0.0433	26.2%
6	Random Forest	0.0560	0.0426	27.2%
6	XGBoost	0.0554	0.0421	28.0%
6	GRU	0.0553	0.0421	28.0%
6	LSTM	0.0561	0.0426	27.1%
12	Persistence	0.0814	0.0618	0.0%
12	Linear Regression	0.0567	0.0433	30.3%
12	Random Forest	0.0563	0.0429	31.0%
12	XGBoost	0.0558	0.0425	31.4%
12	GRU	0.0566	0.0431	30.5%
12	LSTM	0.0562	0.0431	31.0%

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}, \tag{17}$$

$$\text{Skill} = 1 - \frac{\text{RMSE}_{\text{model}}}{\text{RMSE}_{\text{persistence}}}. \tag{18}$$

4 Results

4.1 Model Comparison Across Horizons

Table 6 gives the full six-model comparison at all four horizons (seed= 42). XGBoost and GRU are consistently the top two performers and are essentially tied; skill over persistence rises from about 21% at $h = 1$ to about 31% at $h = 12$, because the persistence baseline itself degrades faster than the learned models as the horizon grows.

Fig. 4 shows RMSE and skill across horizons for all six models, Fig. 5 overlays actual and forecast LQI time series for persistence and XGBoost at $h = 1$ and $h = 12$ h,

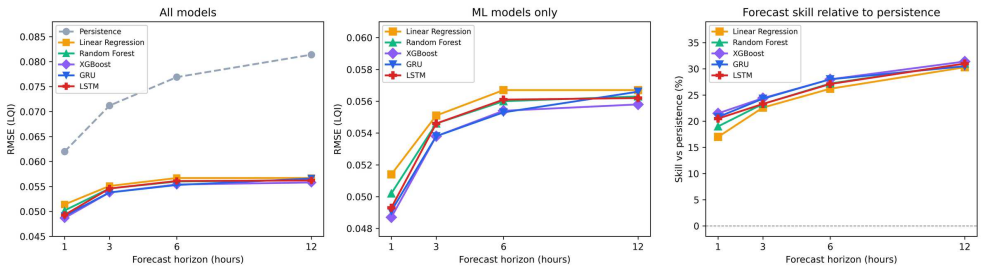

Fig. 4 RMSE and skill vs. persistence across horizons, all six models.

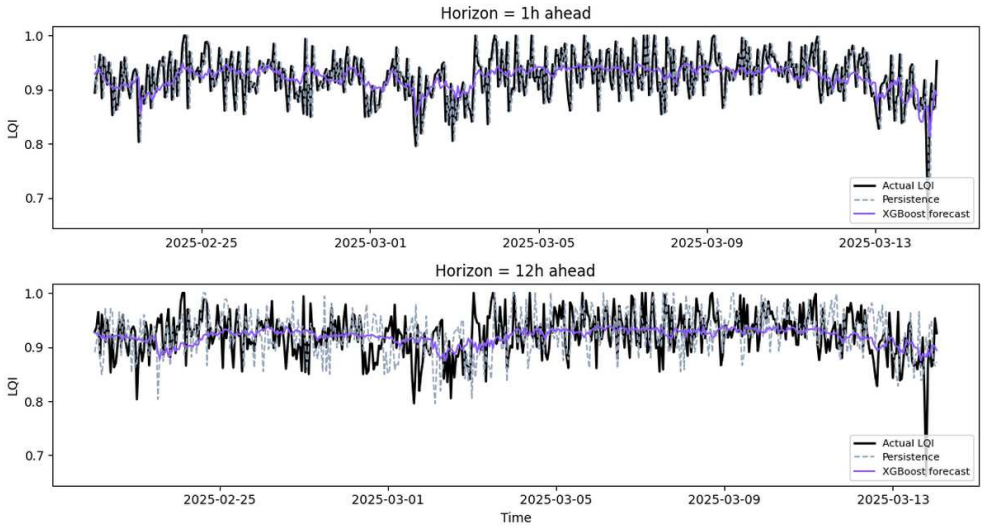

Fig. 5 Actual vs. forecast LQI time series, persistence and XGBoost overlaid, at $h = 1$ h and $h = 12$ h.

and Fig. 8 shows predicted-vs-actual calibration scatter for XGBoost at all four horizons: predictions compress toward the mean at longer horizons and for extreme (very low) LQI values, a limitation revisited directly in the degradation-specific evaluation (Section 4.8).

4.2 Feature Importance and Training Behaviour

Fig. 6 shows XGBoost's global feature importance at $h = 1$ h: `lqi_lag1` dominates by a wide margin, followed by `lqi_lag3`/`lag6`, cyclical day-of-year terms, and rain rate. Fig. 7 confirms GRU/LSTM converge cleanly within a handful of epochs with no train/validation divergence, i.e., no overfitting under the chosen architecture and early-stopping regime.

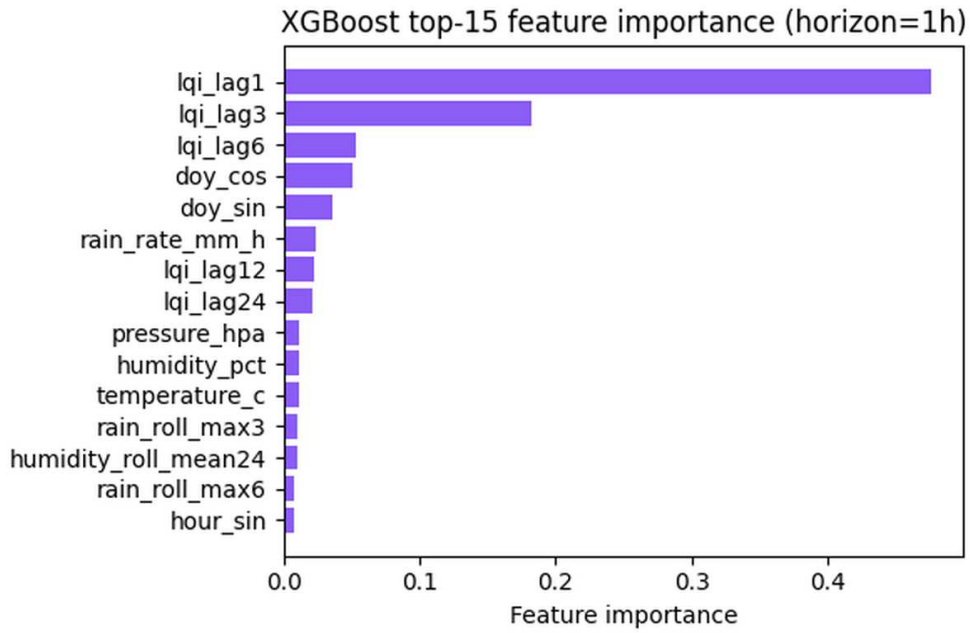

Fig. 6 XGBoost top-15 feature importance, horizon = 1 h.

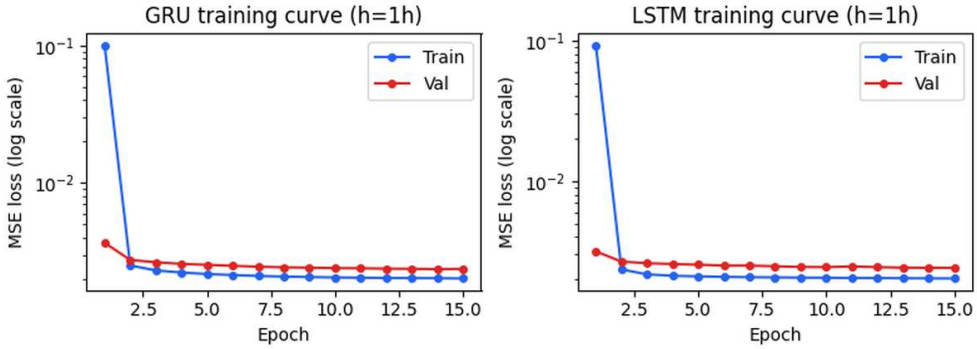

Fig. 7 GRU and LSTM training/validation loss curves (log scale), horizon = 1 h.

4.2.1 An Attention-Based (Transformer) Baseline

A lightweight Transformer encoder [62] (2 layers, $d_{\text{model}} = 32$, 4 attention heads, 17,313 parameters – comparable capacity to the GRU/LSTM baselines) was trained under an identical protocol: the same 24-h lookback window, the same five raw channels, the same TRAIN-only `StandardScaler`, and the same chronological 70/15/15 split, following the general Transformer-for-time-series-forecasting formulation of Lim et al. [40]. No causal mask is needed, since the full lookback window is historical/known at forecast time.

Table 7 Transformer RMSE alongside the Table 6 models (seed= 42).

h (h)	Persist.	XGBoost	GRU	LSTM	Transformer
1	0.0620	0.0487	0.0491	0.0493	0.0490
3	0.0712	0.0538	0.0538	0.0546	0.0532
6	0.0769	0.0554	0.0553	0.0561	0.0554
12	0.0814	0.0558	0.0566	0.0562	0.0574

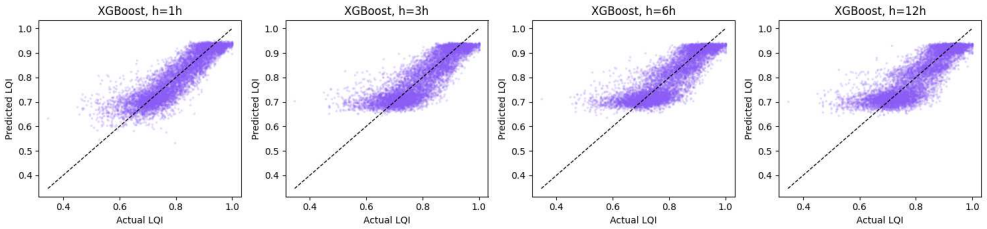

Fig. 8 Predicted-vs-actual calibration scatter for XGBoost at all four horizons.

The Transformer does not meaningfully beat the existing models at any horizon (Table 7): it is roughly tied with XGBoost/GRU/LSTM at $h = 1$–6 h and is noticeably worse at $h = 12$ h (RMSE 0.0574 vs. 0.0558–0.0566). This negative result fits with the ablation finding of Section 4.7: with a 24-h fixed window and roughly 36K training sequences, attention's main advantage (modelling long-range dependencies) has limited room to help, since the model already relies mostly on recent LQI history and current weather rather than distant context.

4.3 Hyperparameter Sensitivity

Every result above used hand-picked defaults (XGBoost depth= 6, learning rate= 0.05; GRU hidden size= 32). A grid search over max_depth × learning_rate for XGBoost, and hidden_size × dropout_p for GRU, was scored on the validation set only, with the test set touched once at the end. This checks whether those defaults were actually near-optimal, following standard hyperparameter-search practice[6].

At $h = 1$ h, the best XGBoost configuration (max_depth= 8, learning rate= 0.05, validation RMSE= 0.0481; final test RMSE= 0.0488) is within 0.0007 RMSE of the hand-picked default (validation RMSE= 0.048135 at depth 6). At $h = 12$ h the best configuration (depth= 8, learning rate= 0.01, validation RMSE= 0.055501) is similarly close to the default (depth= 6, learning rate= 0.05, validation RMSE= 0.055519). For the GRU architecture grid, the best configuration (hidden_size= 32, dropout= 0.0) is exactly the original hand-picked default. The conclusion for both models is that the originally chosen defaults were already close to optimal; the tuning study is reported for methodological completeness not because it meaningfully changes the headline numbers.

Table 8 Multi-seed robustness (5 seeds), XGBoost vs. GRU.

h (h)	Model	Mean RMSE	Std. RMSE	CV%
1	XGBoost	0.0507	0.0002	0.3%
1	GRU	0.0518	0.0004	0.7%
12	XGBoost	0.0590	0.0001	0.2%
12	GRU	0.0602	0.0002	0.3%

Table 9 Diebold–Mariano test, XGBoost vs. GRU, all four horizons.

h (h)	XGB RMSE	GRU RMSE	DM stat.	p-value	Sig. (5%)
1	0.0487	0.0491	−1.89	0.058	No
3	0.0538	0.0538	−0.22	0.827	No
6	0.0554	0.0553	+0.13	0.897	No
12	0.0558	0.0566	−2.30	0.021	Yes

4.4 Multi-Seed Robustness

Every result above used a single seed (42). Table 8 reruns XGBoost and GRU across five seeds (model-internal randomness only; the chronological split itself is never reshuffled) at horizons 1 h and 12 h.

At both horizons the XGBoost-GRU gap (∼0.0012 RMSE) exceeds the combined standard deviation of the two models, suggesting the gap is likely real rather than run-to-run noise, giving statistical backing to the claim that XGBoost is the stronger classical/sequence-model comparator at this scale of data. (Absolute RMSE values differ slightly from Table 6 because that table used seed= 42 specifically, while Table 8 averages seeds 1–5; the ordering and conclusion are unaffected.)

4.4.1 Point-Wise Significance: Diebold–Mariano Test

Multi-seed robustness asks whether the XGBoost-vs-GRU ranking holds under re-initialization; it does not test whether the two models' forecast errors are distinguishable on a point-by-point basis for a given trained pair. The Diebold–Mariano test [15] answers exactly that: for each horizon, the two models' squared forecast errors are compared on the identical set of test timestamps (seed= 42), using a loss differential $d_t = e^2_{\mathrm{XGB},t} - e^2_{\mathrm{GRU},t}$ under H_0: equal predictive accuracy. Variance is estimated with a Newey–West (HAC) correction using a truncation lag of $h-1$, following Diebold and Mariano's own recommendation for h-step-ahead forecasts, together with the Harvey–Leybourne–Newbold small-sample correction [22].

The XGBoost-GRU gap is statistically significant only at $h = 12$ h (DM= -2.30, $p = 0.021$); at $h = 1, 3$, and 6 h the two models are statistically indistinguishable (all $p > 0.05$), despite the multi-seed test above suggesting the gap was likely real at both horizons it checked. Table 9 reports the full test statistics and p-values at all four

Table 10 MC Dropout coverage (target: ∼95%).

h (h)	RMSE	MAE	R^2	95% CI Coverage
1	0.0507	0.0388	0.7660	92.9%
3	0.0555	0.0428	0.7189	90.0%
6	0.0563	0.0430	0.7111	90.3%
12	0.0571	0.0439	0.7031	89.8%

horizons. This sharpens, rather than contradicts, the multi-seed finding: the two tests ask different questions (re-initialization robustness vs. point-wise error significance), and the precise, honest claim is that XGBoost and GRU perform comparably at short-to-medium horizons, with XGBoost's advantage only becoming statistically confirmed at the longest horizon tested.

4.5 Predictive Uncertainty: Monte Carlo Dropout

An explicit dropout layer [58] ($p = 0.2$) is added before the GRU's output head, necessary because PyTorch's built-in RNN dropout is a silent no-op for a single-layer recurrent module. At inference, dropout is kept active and $T = 100$ stochastic forward passes $f_t(x)$ are run per test point [18], giving a predictive mean and standard deviation:

$$\hat{\mu}(x) = \frac{1}{T}\sum_{t=1}^{T} f_t(x), \tag{19}$$

$$\hat{\sigma}^2(x) = \frac{1}{T}\sum_{t=1}^{T}(f_t(x) - \hat{\mu}(x))^2, \tag{20}$$

from which the approximate 95% predictive interval follows directly, $\mathrm{CI}_{95\%} = \hat{\mu} \pm 1.96\hat{\sigma}$. Table 10 reports coverage (target: ∼95%) at all four horizons; ensemble-based alternatives [38] and broader uncertainty-quantification approaches [23] are discussed as future directions in Section 5.3. Fig. 9 illustrates this behavior for two representative horizons, plotting the 95% uncertainty band together with the actual and mean forecast LQI.

Fig. 10 shows the more informative, honest finding on its right-hand panel: 95% CI coverage drops from ∼93% to ∼90% as horizon grows, meaning the model becomes progressively *overconfident* even as RMSE visibly rises over the same horizons – a genuine, reportable limitation for any risk-aware downstream controller.

4.6 Explainability: SHAP

SHAP (TreeExplainer) [43, 44] attributes each XGBoost prediction to individual feature contributions, going beyond the global importance ranking of Fig. 6 to show direction and per-prediction attribution; LIME [53] and broader explainable-AI frameworks [5] offer complementary perspectives not pursued here. Fig. 11 (beeswarm,

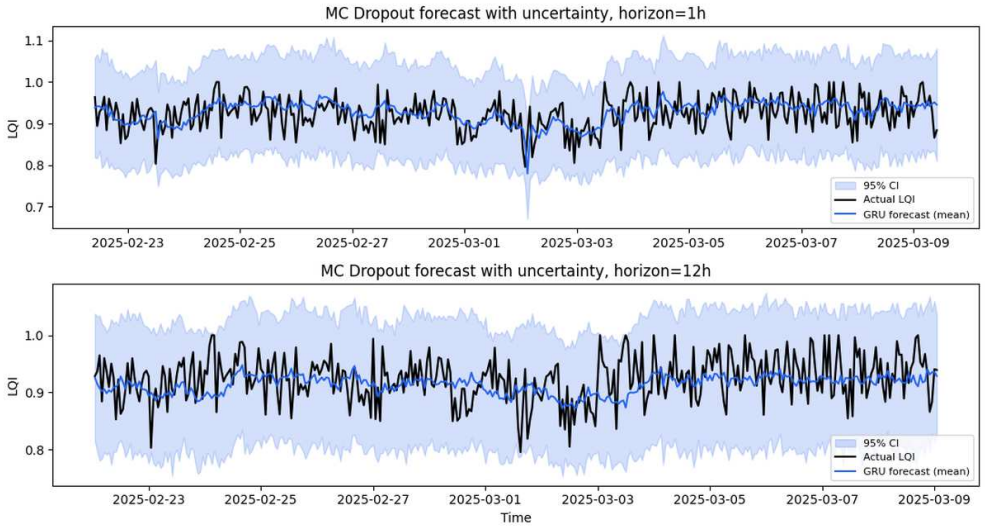

Fig. 9 MC Dropout forecast with 95% uncertainty band, actual LQI overlaid, at $h = 1$ h and $h = 12$ h.

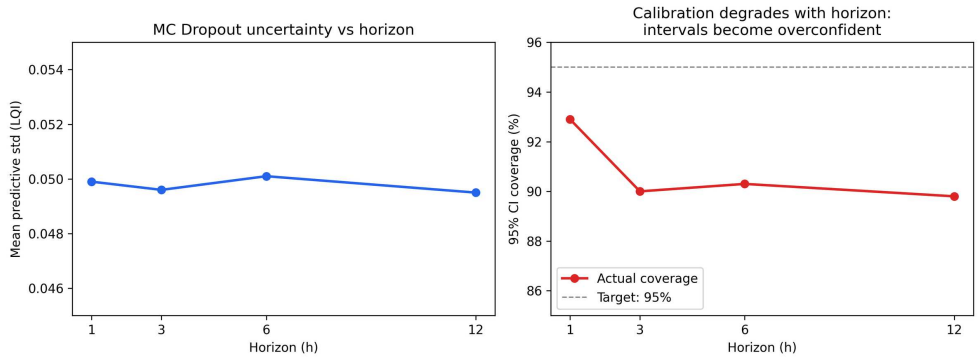

Fig. 10 Left: predictive std. stays roughly flat across horizons. Right: 95% CI coverage drops from ∼93% to ∼90% as horizon grows – the model becomes progressively overconfident.

$h = 1$ h, $n = 3000$ sampled test points) shows `lqi_lag1` as the dominant driver in both magnitude and consistency, with humidity, temperature, and rain rate contributing meaningfully in the expected directions. Fig. 12 shows a single-prediction waterfall for a high-rain example (rain= 3.3 mm/h), decomposing the forecast from the baseline expectation $E[f(X)] = 0.822$ down to the final prediction $f(x) = 0.654$; rain rate and `lqi_lag1` are the two largest negative contributors for this degraded-conditions forecast.

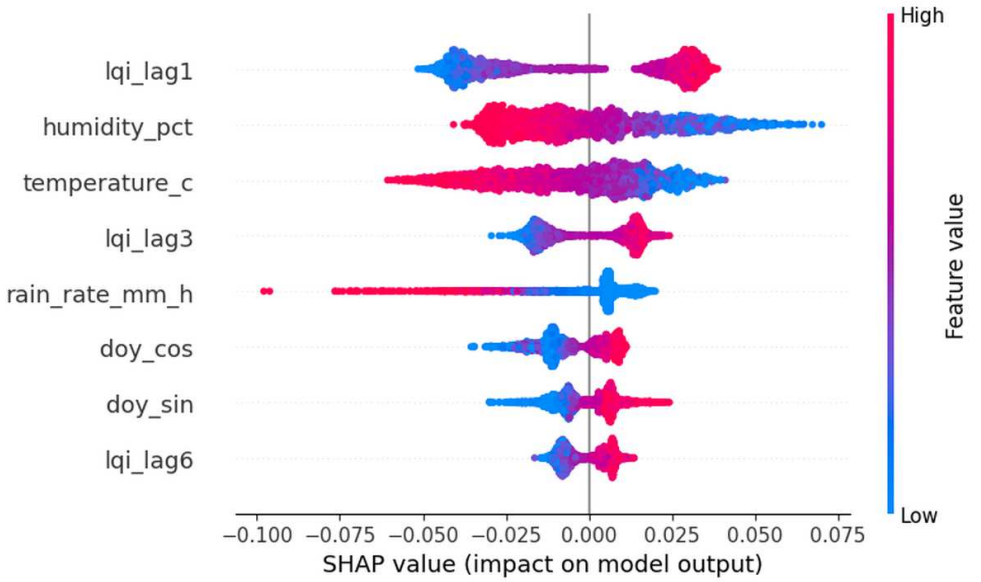

Fig. 11 SHAP summary (beeswarm) plot – global feature impact and directionality across 3,000 sampled test predictions, horizon = 1 h.

Table 11 Skill vs. persistence (%) by feature set and horizon.

h (h)	S1 weather_raw	S2 weather_hist.	S3 lqi_hist.	S4 full
1	+22.4%	+21.6%	+21.9%	+21.6%
3	+25.0%	+24.6%	+24.7%	+24.4%
6	+27.4%	+28.2%	+27.4%	+28.0%
12	+30.0%	+31.0%	+31.0%	+31.4%

4.7 Feature-Group Ablation

SHAP explains a single trained model's internal attributions but does not answer a different question: what happens if a feature group had never been given to the model at all? Four nested XGBoost feature sets isolate this: S1 *weather_raw* (instantaneous weather and time only, $n = 8$), S2 *weather_history* (S1 + rain/humidity lags and rolling statistics, $n = 22$), S3 *lqi_history* (S1 + LQI lags only, $n = 13$), and S4 *full* (all combined, $n = 27$, matching the main pipeline).

Remarkably, S1 (instantaneous weather alone, only 8 features, no history of any kind) already captures most of the achievable skill at every horizon – adding weather history (S2) or LQI autocorrelation (S3) individually, or both together (S4), yields only marginal further gains (roughly 0–1.5 percentage points). The model is not simply riding LQI's own autocorrelation like a disguised persistence model (S3 barely beats S1), and engineered history features add real but modest value over the current

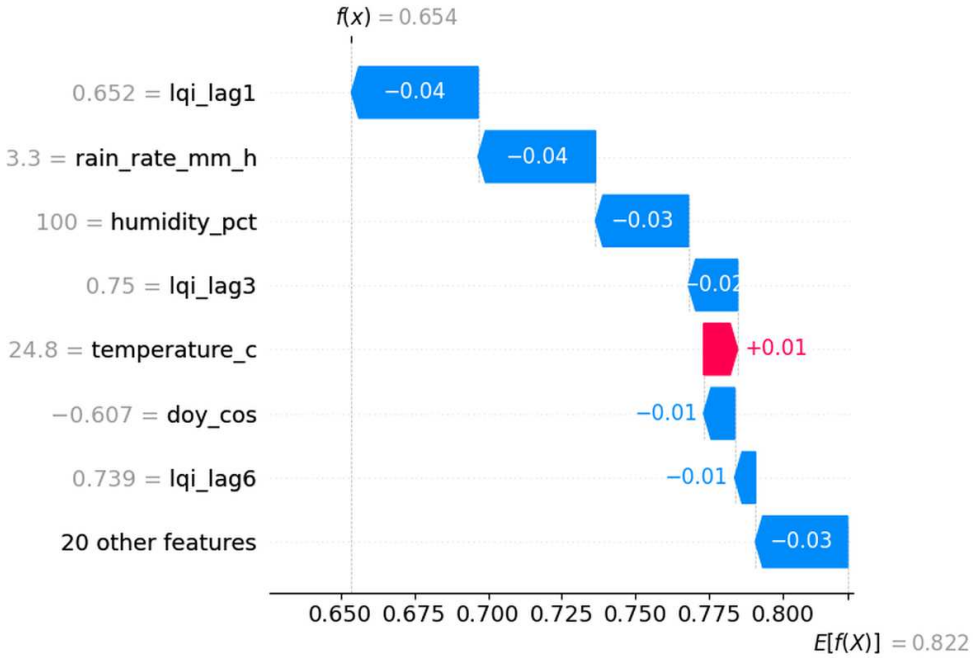

Fig. 12 SHAP waterfall for one individual high-rain-rate test prediction (rain= 3.3 mm/h), horizon = 1 h.

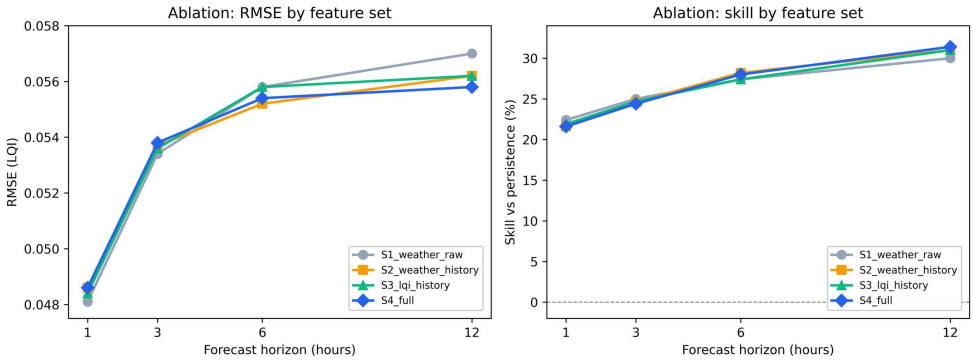

Fig. 13 Ablation study: RMSE and skill-vs-persistence by nested feature set, across all four horizons.

instantaneous weather reading – a pattern also observed in weather-driven renewable-energy forecasting, where instantaneous meteorological inputs likewise capture most of the achievable skill [4, 14]. Fig. 13 visualizes this pattern across all four horizons, plotting both RMSE and skill-vs-persistence by nested feature set.

This finding raises a fair question for a venue focused on advanced signal processing and learning methods: if instantaneous weather alone is nearly sufficient, why pursue

Table 12 Stratified regression metrics, horizon = 1 h.

Stratum	n	RMSE	R^2	Skill
Overall	7,780	0.0487	0.784	21.5%
No rain at target	5,081	0.0402	0.765	24.3%
Rain at target	2,699	0.0615	0.417	19.0%
Worst LQI quartile	1,945	0.0643	−0.591	18.7%
Event window (±2 h)	4,034	0.0567	0.566	19.8%
Outside event window	3,746	0.0382	0.753	25.1%

history-aware sequence models and deep learning at all? A few points are worth making explicitly.

The marginal value of history is not static. It is essentially flat-to-negative at $h = 1$ (S1 +22.4% vs. S4 +21.6%) but grows to a consistent +1.4 percentage-point advantage for the full feature set by $h = 12$ (S1 +30.0% vs. S4 +31.4%, Table 11), suggesting that history and sequence modeling become *more* rather than less valuable as the forecasting task hardens: a trend that would plausibly continue beyond the 12-hour horizon tested here.

There is also value in the negative result itself. Establishing that a simple, cheap, interpretable model is sufficient at this problem scale is an operationally useful finding, not a reason to abandon the comparison: it directly informs deployment choice for resource-constrained outdoor THz hardware, where a lightweight XGBoost model is preferable to a recurrent network of comparable accuracy.

The recurrent and attention-based architectures earn their place for a different reason, they natively support the sequence-based Monte Carlo Dropout uncertainty quantification of Section 4, which a single-point gradient-boosted regressor does not provide without additional machinery, not for a marginal RMSE edge. A key limitation of single-site data is that it caps how much value weather history can add beyond the current reading, since rain events at one location tend to follow similar temporal patterns. To test whether learned temporal dynamics genuinely outperform instantaneous snapshots, Section 5.2 extends the analysis to a second, climatically different site (Rajshahi), precisely the setting where such dynamics should matter most.

4.8 Degradation-Specific Evaluation

Every metric so far is computed over the entire test set. But 68% of this dataset has zero rain and a mean LQI of 0.82, i.e., the test set is dominated by benign, easy-to-forecast conditions. A model can post an excellent aggregate RMSE while forecasting badly during exactly the rain events that motivate this research question, and the aggregate number would never reveal it: a general risk noted in weather-driven and precipitation-forecasting evaluation [52, 56]. Table 12 stratifies evaluation by condition at $h = 1$ h (representative; all four horizons follow the same pattern). Fig. 14 visualizes this stratified RMSE pattern by condition across all four horizons.

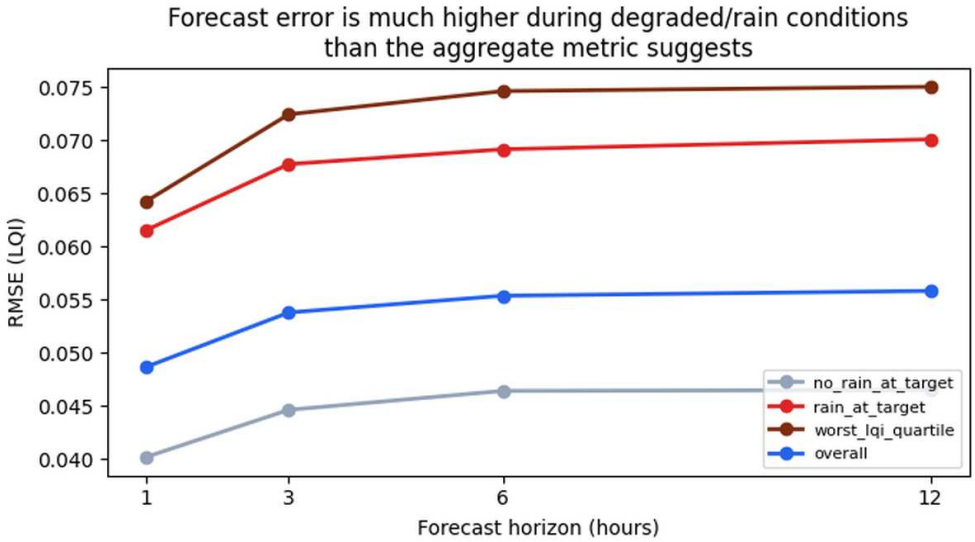

Fig. 14 Stratified RMSE by condition (no-rain / rain / worst-LQI-quartile / overall) across all four horizons.

Table 13 Binary degradation detection: LQI $<$ 25th-percentile(train).

h (h)	Model P/R/F1	Persist. P/R/F1
1	0.741 / 0.829 / 0.782	0.721 / 0.721 / 0.721
3	0.680 / 0.813 / 0.741	0.655 / 0.655 / 0.655
6	0.667 / 0.810 / 0.731	0.616 / 0.616 / 0.616
12	0.669 / 0.804 / 0.730	0.571 / 0.571 / 0.571

The worst-LQI-quartile R^2 is negative at $h = 1$ h (-0.591) and gets worse at longer horizons (down to about -1.17 at $h = 12$ h), meaning in exactly the conditions this research is meant to serve, the model performs worse than simply predicting the mean of that stratum. This is the single most important honest finding in the whole pipeline.

Reframing the regression forecast as a binary question: will LQI at $t+h$ drop below a degradation threshold, fixed at the 25th percentile of TRAINING LQI (= 0.7345, leak-free), is arguably the more operationally relevant framing for a downstream controller, since it only needs an early warning, not the exact LQI value. Table 13 reports precision/recall/F1 for the model versus persistence. Fig. 15 plots precision, recall, and F1 against horizon for both the model and the persistence baseline, and Fig. 16 shows the corresponding confusion matrices at the shortest ($h = 1$ h) and longest ($h = 12$ h) horizons, illustrating the precision/recall trade-off underlying these aggregate scores.

The model beats persistence on F1 at every horizon, and the gap widens with horizon (persistence F1 falls from 0.721 to 0.571; model F1 stays around 0.73–0.78). This is an encouraging counterpoint to the regression-stratified result above: even

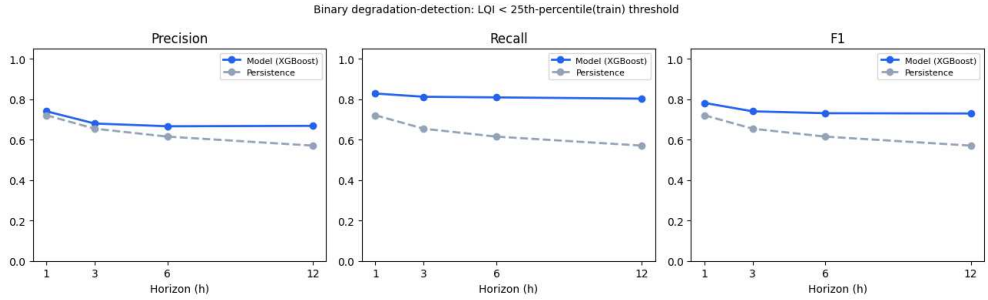

Fig. 15 Degradation-detection precision, recall, and F1 vs. horizon, model vs. persistence.

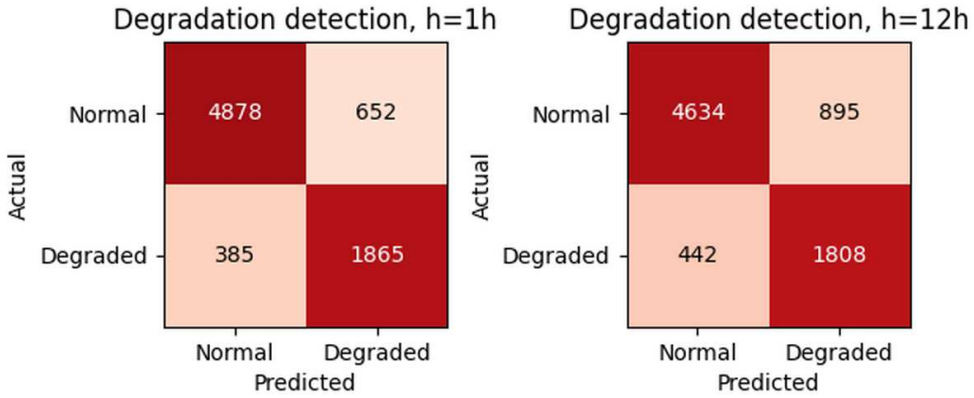

Fig. 16 Confusion matrices for degradation detection at $h = 1$ h and $h = 12$ h.

though exact-value forecasting degrades badly in worst-quartile conditions, the binary early-warning signal that a downstream controller would actually consume remains meaningfully better than the naive baseline throughout. Both results belong in this paper; together they give a fuller, more honest picture than either alone.

5 Discussion

5.1 Key Findings Summary

Forecast skill grows steadily with horizon, from 21.5% at $h = 1$ to 31.4% at $h = 12$, mainly because persistence degrades faster than any of the learned models as the window lengthens. XGBoost and GRU track each other closely at short and medium horizons (1–6 h), and the Diebold–Mariano test confirms they are statistically indistinguishable there (all $p > 0.05$). Only at $h = 12$ does XGBoost pull ahead with statistical backing (DM$= -2.30$, $p = 0.021$), and it does so while being far cheaper to train and tune. Adding a Transformer baseline does not change this picture: it fails to beat the simpler models at any horizon, which fits with what the ablation study already suggests: instantaneous weather alone gets most of the way there. Hand-picked

hyperparameters also turned out to be close enough to optimal that a full grid search barely moved the numbers; we ran it anyway, mainly for methodological completeness.

Where the picture gets less comfortable is uncertainty and worst-case behaviour. MC Dropout intervals are reasonably well calibrated at short horizons but grow overconfident as the horizon extends, which matters for any downstream controller trying to act on these forecasts. SHAP and the ablation study agree on the same underlying story: LQI's own recent history (`lqi_lag1` especially) carries the most signal, though instantaneous weather alone already captures most of the achievable skill. The most consequential result, however, is not a positive one: aggregate metrics look fine, but they hide a real collapse in performance, with regression skill turning R^2-negative in exactly the worst-LQI-quartile and rain conditions this work is meant to address: even though reframing the problem as a binary degradation alarm rather than an exact-value forecast still beats persistence comfortably in those same conditions.

The ablation results raise a fair question about the role of deep learning in this pipeline: instantaneous weather alone (S1, 8 features) already captures most of the achievable skill, and GRU, LSTM, and the Transformer all fail to beat XGBoost by any practically meaningful margin. The recurrent and attention-based models are nonetheless retained here for a few reasons.

First, they establish a fair comparison. Without GRU, LSTM, and Transformer baselines run under an identical protocol, the claim that simple features are enough would remain an untested assumption rather than a demonstrated finding.

Second, the sequence models are also not dead weight. MC Dropout uncertainty quantification needs a differentiable, stochastic model with an active dropout layer at inference, and XGBoost does not offer that natively. The GRU is what makes the calibrated predictive intervals possible at all, independent of whether its point forecasts beat XGBoost's.

There is a site-specific angle worth flagging too. Sylhet's test period is 68% rain-free (Section 4.8), so recent LQI history and current weather may simply dominate here because sustained, multi-hour degradation episodes do not come up often in this particular sample. A site with more persistent frontal rain, or a horizon longer than the 12 h tested here, could give a sequence model's temporal modelling more room to matter, and the recurrent architectures built here are a reasonable starting point for testing that rather than something to discard.

Finally, XGBoost's own top feature, `lqi_lag1`, is itself a one-step temporal dependency. That a tree exploiting a single lag can match a recurrent network working with a full 24-hour window says something about the temporal structure of this particular dataset; it does not mean temporal modelling is pointless in general.

Taken together, the fair conclusion is not that deep learning was wasted effort, but something more specific: for this site, this climate, and this range of horizons, a well-tuned gradient-boosted tree is a cheaper, sufficient point-forecaster, while the sequence-model machinery earns its place through uncertainty quantification and as groundwork for testing whether temporal modelling pays off elsewhere.

Table 14 Cross-site comparison: XGBoost forecast skill vs. persistence (%), by horizon.

h (h)	Sylhet	Rajshahi
1	21.5%	23.0%
3	24.4%	25.0%
6	28.0%	28.1%
12	31.4%	32.5%

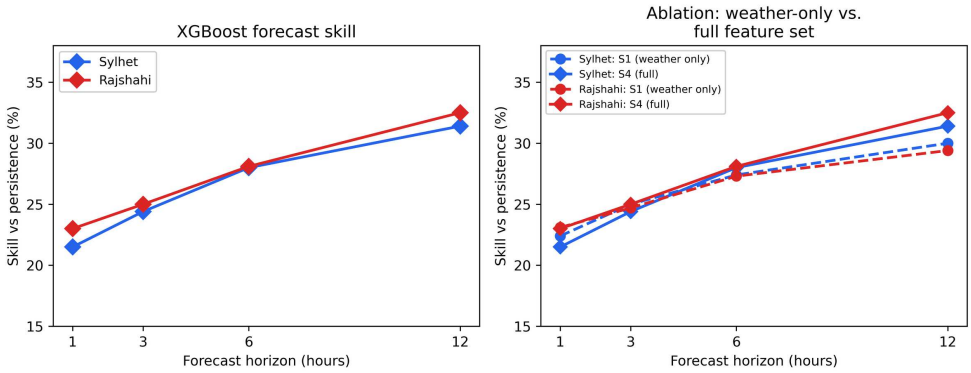

Fig. 17 Cross-site comparison. Left: XGBoost forecast skill vs. persistence, Sylhet vs. Rajshahi. Right: the ablation finding (instantaneous weather alone, S1, vs. the full feature set, S4) replicated at both sites.

5.2 Cross-Site Validation: Sylhet vs. Rajshahi

Every result reported so far comes from a single site (Sylhet). To check whether these findings are a Sylhet-specific artefact or hold more generally, we re-ran the identical physics simulator, feature engineering, and seven-model benchmark (Sections 2–3, unchanged) on an independent six-year weather record for Rajshahi, Bangladesh (24.37°N, 88.60°E), a north-western district with a noticeably drier climate than Sylhet, covering the identical date range (2020-01-01 to 2025-12-01, 51,888 hourly rows) for a fair comparison. No model architecture, hyperparameter, or evaluation protocol changed between sites; only the weather record did. Rain rate is far more strongly coupled to LQI at Sylhet than at Rajshahi ($\mathrm{corr}(\mathrm{LQI}, \mathrm{rain}) = -0.490$ vs. -0.358), which fits Rajshahi's drier profile, while both sites show comparably strong coupling to total attenuation (-0.925 vs. -0.924) – a sign that the physics simulator behaves consistently across climates.

The central findings hold up. XGBoost remains the best or joint-best model at every horizon at both sites, and the two sites land on closely comparable forecast skill (Table 14, Fig. 17): if anything, Rajshahi comes out slightly ahead despite its weaker rain–LQI coupling. The ablation pattern repeats as well: instantaneous weather alone (S1) captures most of the achievable skill at both sites, and the advantage of the full feature set (S4) is flat or slightly negative at $h = 1$ (Sylhet: +22.4% vs. +21.6%;

Rajshahi: +23.1% vs. +22.7%) before growing into a real, comparably sized edge by $h = 12$ (Sylhet: +30.0% vs. +31.4%; Rajshahi: +29.4% vs. +32.5%). Even the SHAP rankings agree: `lqi_lag1`, `humidity_pct`, and `temperature_c` top the list at $h = 1$ at both sites.

So does the weakest result – and it is somewhat worse at Rajshahi. The worst-LQI-quartile stratum is R^2-negative at both sites, and gets worse with horizon: -0.591 at $h = 1$ rising to -1.169 at $h = 12$ for Sylhet, versus -0.711 and -1.229 for Rajshahi. Binary degradation detection still beats persistence at both sites (F1 at $h = 12$: 0.730 vs. 0.571 for Sylhet; 0.665 vs. 0.530 for Rajshahi). If there is one piece of evidence in this paper that the worst-case forecast failure is a property of the LQI-forecasting problem itself, rather than something specific to Sylhet's data, this is it.

Two secondary results do not carry over as cleanly, and we report that plainly rather than quietly setting them aside. The Diebold–Mariano significance at $h = 12$ is site-dependent: XGBoost's edge over GRU is significant at Sylhet (DM= -2.30, $p = 0.021$) but not at Rajshahi (DM= -1.40, $p = 0.161$), and the multi-seed check agrees – the gap is "likely real" at Sylhet but "within noise" at Rajshahi. The Transformer's long-horizon weakness is also site-dependent: noticeably worse than the other models at $h = 12$ for Sylhet (RMSE 0.0574 vs. 0.0558–0.0566) but fully competitive at Rajshahi (RMSE 0.0528, tied with LSTM). Both discrepancies point the same way: XGBoost's specific edge at the longest horizon is real at Sylhet but not something we can promise elsewhere, while the broader claims – that simple weather features carry most of the skill, and that aggregate metrics hide serious degraded-condition failure – hold up across both climates.

Two smaller differences are worth mentioning as context rather than concern. MC Dropout calibration is actually better at Rajshahi (95% CI coverage starts at 95.7% and only falls to 93.3% by $h = 12$) than at Sylhet (92.9% falling to 89.8%), plausibly because lower rain variance makes the forecasting problem less chaotic there. The best XGBoost tree at Rajshahi is also shallower and learns faster (depth= 4, learning rate= 0.10) than at Sylhet (depth= 8, learning rate= 0.05), consistent with a simpler, less rain-driven LQI structure at the drier site.

5.3 Limitations

The simulator is physics-model-derived, driven by real weather observations, but is not a direct THz field measurement, consistent with the broader literature's reliance on channel measurement campaigns for full validation [2, 41, 45, 51, 61, 64]; this is stated plainly rather than implied. Weather data resolution (hourly) may be coarser than ideal for rain-rate-sensitive THz attenuation. Approximately 11% of LQI variance is attributable to a weather-independent hardware/pointing-loss term by design, so R^2 against weather-only LQI is worth reporting as an honest ceiling metric in future extensions of this work. MC Dropout's predictive standard deviation stays roughly flat across horizons rather than growing with genuine forecast difficulty, likely reflecting model/architecture uncertainty more than data-driven uncertainty; ensemble-based [38] or conformal-prediction approaches to uncertainty quantification [23] are a direction for future refinement rather than a flaw to hide. Finally, while Section 5.2 shows

the central findings replicate at a second, climatically contrasting site (Rajshahi), both sites are located in Bangladesh; broader geographic and climatic generalization (e.g., temperate, arid, or maritime tropical regions), consistent with the scope of prior Bangladesh-specific rain-attenuation studies at other frequency bands [25, 26], remains future work.

5.4 Toward Field Validation: A Concrete Roadmap

The absence of a THz field measurement campaign is the most consequential limitation of this study, and it is worth being concrete about how the present framework is intended to be validated rather than leaving this as an open-ended remark. The physics-based simulator itself rests on internationally standardized, independently validated propagation models (ITU-R P.838-3 and P.676) rather than on an ad hoc channel assumption, so the forecasting layer built on top of it is validated against a physically grounded, if not field-measured, channel. The natural next experiment would be a short-range (tens to a few hundred metres), fixed 300 GHz link with a commercial or university-hosted THz transceiver pair, co-located with an automatic weather station, allowing direct comparison of measured received power and BER against the simulator's predictions under real rain events; this requires hardware access beyond the scope of the present simulation-based study, but not new theory, only instrumentation and time. Before such a campaign is feasible, the framework can be partially cross-validated against publicly available THz/sub-THz outdoor channel measurement datasets reported elsewhere in the literature [2, 51, 61], by checking whether the fitted ITU-R coefficients and attenuation ranges used here fall within the envelopes those campaigns report; this literature cross-check is flagged as a concrete, low-cost next step. The multi-site extension identified as the highest-priority action in an earlier version of this roadmap has, in fact, already been carried out: Section 5.2 re-runs the identical pipeline on Rajshahi, a climatically contrasting site, and finds that the central conclusions replicate. The natural continuation is to extend this same near-zero-cost approach beyond Bangladesh, to sites with genuinely different climate regimes such as temperate, arid, or maritime ones, which would test the boundaries of generalization more aggressively than two same-country sites can. These are reported as concrete, prioritized next steps precisely so that the absence of field data is read as a deliberate, resource-bounded scoping decision for the present contribution rather than an oversight. In parallel, the trained forecasting engine will be extended to drive predictive adaptive-modulation and reconfiguration strategies, benchmarked against reactive baselines, drawing on reinforcement-learning-based resource allocation approaches recently explored for adjacent wireless domains [46].

6 Conclusion

This work puts together a complete, methodologically careful weather-aware LQI forecasting pipeline for outdoor THz links: a physics-based, standards-compliant, leak-free simulator; a composite, physically anchored Link Quality Index with a transparently corrected flaw; a seven-model forecasting benchmark, including an attention-based Transformer, backed by hyperparameter tuning, multi-seed robustness checks, and

point-wise significance testing; uncertainty quantification and explainability; and, most importantly, a degradation-conditioned evaluation that shows exactly where and by how much the aggregate numbers overstate real-world performance. Section 5.2 shows that these findings, including the most consequential one that aggregate metrics hide severe worst-case failure, hold up on an independent, climatically different site (Rajshahi), suggesting they are properties of the LQI-forecasting problem rather than quirks of one location's data. Building on the roadmap in Section 5.4, the next steps in order of priority are extending the cross-site evaluation beyond Bangladesh to climate regimes that test generalization more aggressively, checking the simulator's fitted attenuation ranges against published outdoor THz/sub-THz measurement campaigns as an interim, literature-based validation step, and pursuing a short-range fixed THz link measurement campaign, co-located with weather instrumentation, to directly validate the simulator against field data as resources allow.

References

[1] Akyildiz IF, Han C, Hu Z, et al (2022) Terahertz band communication: An old problem revisited and research directions for the next decade. IEEE Transactions on Communications 70:4250–4285. https://doi.org/10.1109/TCOMM.2022.3171800

[2] Al-jzari A, He Y, Hu J, et al (2025) Sub-THz and THz channel measurements and characteristic analysis in indoor and outdoor environments for 6G wireless systems. IEEE Transactions on Terahertz Science and Technology 15:976–984. https://doi.org/10.1109/TTHZ.2025.3598719

[3] Albelali S, Ahmed M (2025) Hidden leaks in time series forecasting: How data leakage affects LSTM evaluation across configurations and validation strategies. arXiv:2512.06932

[4] Aldawsari MA, Hassan BM, et al (2026) Enhancing wind and solar energy forecasting through time-series feature engineering and ensemble machine learning. Scientific Reports 16:15546. https://doi.org/10.1038/s41598-026-49373-7

[5] Arrieta AB, Díaz-Rodríguez N, Del Ser J, et al (2020) Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58:82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[6] Bergstra J, Bengio Y (2012) Random search for hyper-parameter optimization. Journal of Machine Learning Research 13:281–305. https://jmlr.org/papers/v13/bergstra12a.html

[7] Breiman L (1996) Bagging predictors. Machine Learning 24:123–140. https://doi.org/10.1007/BF00058655

[8] Breiman L (2001) Random forests. Machine Learning 45:5–32. https://doi.org/10.1023/A:1010933404324

[9] Cai Y, et al (2024) Photonics-aided exceeding 200-Gb/s wireless data transmission over outdoor long-range 2×2 MIMO THz links at 300 GHz. Optics Express 32:33587–33602. https://doi.org/10.1364/OE.534985

[10] Cano J, Israël J, Féral L (2025) AIRIS2: A smart gateway diversity algorithm for very high-throughput satellite systems. IEEE Wireless Communications Letters 14(7):2189–2193. https://doi.org/10.1109/LWC.2025.3565705

[11] Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. In: Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), https://doi.org/10.1145/2939672.2939785

[12] Chowdhury MZ, Shahjalal M, Ahmed S, et al (2020) 6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions. IEEE Open Journal of the Communications Society 1:957–975. https://doi.org/10.1109/OJCOMS.2020.3010270

[13] Chung J, Gulcehre C, Cho K, et al (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. In: NIPS 2014 Workshop on Deep Learning, arXiv:1412.3555

[14] Dhungana H (2025) A machine learning approach for wind turbine power forecasting for maintenance planning. Energy Informatics 8:2. https://doi.org/10.1186/s42162-024-00459-4

[15] Diebold FX, Mariano RS (1995) Comparing predictive accuracy. Journal of Business & Economic Statistics 13:253–263. https://doi.org/10.1080/07350015.1995.10524599

[16] Ferdowsi A, Whitefield D (2021) Deep learning for rain fade prediction in satellite communications. In: Proc. IEEE Global Communications Conference (GLOBECOM), https://ieeexplore.ieee.org/document/9682090

[17] Friedman JH (2001) Greedy function approximation: A gradient boosting machine. The Annals of Statistics 29:1189–1232. https://doi.org/10.1214/aos/1013203451

[18] Gal Y, Ghahramani Z (2016) Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: Proc. 33rd Int. Conf. Machine Learning (ICML), https://proceedings.mlr.press/v48/gal16.html

[19] Giordani M, Polese M, Mezzavilla M, et al (2020) Toward 6G networks: Use cases and technologies. IEEE Communications Magazine 58:55–61. https://doi.org/10.1109/MCOM.001.1900411

[20] Gizzini AK, Medjahdi Y, Ghandour AJ, et al (2025) Explainable AI for enhancing efficiency of DL-based channel estimation. IEEE Transactions on Machine Learning in Communications and Networking 3:976–996. https://ieeexplore.ieee.org/document/11115091

[21] Han C, Wang Y, Li Y, et al (2022) Terahertz wireless channels: A holistic survey on measurement, modeling, and analysis. IEEE Communications Surveys & Tutorials 24:1670–1707. https://doi.org/10.1109/COMST.2022.3182539

[22] Harvey D, Leybourne S, Newbold P (1997) Testing the equality of prediction mean squared errors. International Journal of Forecasting 13:281–291. https://doi.org/10.1016/S0169-2070(96)00719-4

[23] He W, Jiang Z, Xiao T, et al (2026) A survey on uncertainty quantification methods for deep learning. ACM Computing Surveys 58(7):179:1–179:35. https://doi.org/10.1145/3786319

[24] Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Computation 9:1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[25] Hossain S (2014) Rain attenuation prediction for terrestrial microwave link in Bangladesh. Journal of Electrical and Electronic Engineering 7:63–68

[26] Hossain S, Islam MA (2017) Estimation of rain attenuation at EHF bands for earth-to-satellite links in Bangladesh. In: Proc. Int. Conf. Electrical, Computer and Communication Engineering (ECCE), IEEE, pp 589–593, https://ieeexplore.ieee.org/document/7912973

[27] International Telecommunication Union (2005) Recommendation ITU-R P.838-3: Specific attenuation model for rain for use in prediction methods. Tech. rep., ITU-R, https://www.itu.int/rec/R-REC-P.838/en

[28] International Telecommunication Union (2015) IMT vision – framework and overall objectives of the future development of IMT for 2020 and beyond. Tech. Rep. Recommendation ITU-R M.2083, ITU-R, Geneva, Switzerland, https://www.itu.int/rec/R-REC-M.2083/en

[29] International Telecommunication Union (2022) Recommendation ITU-R P.676-13: Attenuation by atmospheric gases and related effects. Tech. rep., ITU-R, https://www.itu.int/rec/R-REC-P.676/en

[30] Itahara S, Nishio T, Morikura M, et al (2020) Online trainable wireless link quality prediction system using camera imagery. In: Proc. IEEE Global Commun. Conf. Workshops (GLOBECOM Wkshps), pp 1–6, https://doi.org/10.1109/GCWkshps50303.2020.9367396

[31] Jacoby D, Ostrometzky J, Messer H (2020) Short-term prediction of the attenuation in a commercial microwave link using LSTM-based RNN. In: Proc. 28th European Signal Processing Conference (EUSIPCO), pp 1628–1632, https://doi.org/10.23919/Eusipco47968.2020.9287835

[32] Jahan I, Islam MJ (2025) Impact of rain attenuation on the performance of 6g terahertz communication system. In: 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN), Rangpur, Bangladesh, https://doi.org/10.1109/QPAIN66474.2025.11171880

[33] Jiang W, Zhou Q, He J, et al (2024) Terahertz communications and sensing for 6G and beyond: A comprehensive review. IEEE Communications Surveys & Tutorials 26:2326–2381. https://doi.org/10.1109/COMST.2024.3385908

[34] Kanto Y, Watabe K (2024) Wireless link quality estimation using LSTM model. In: Proc. IEEE/IFIP Network Operations and Management Symposium (NOMS), Seoul, South Korea, pp 1–5, https://ieeexplore.ieee.org/document/10575638

[35] Khan N, Abdallah A, Celik A, et al (2025) Explainable AI-aided feature selection and model reduction for DRL-based V2X resource allocation. IEEE Transactions on Communications https://doi.org/10.1109/TCOMM.2025.3554655

[36] Kingma DP, Ba J (2015) Adam: A method for stochastic optimization. In: Proc. Int. Conf. Learning Representations (ICLR), arXiv:1412.6980

[37] Kouhalvandi L, Matekovits L (2025) On the role of artificial intelligent technology for millimetre-wave and terahertz applications. Sensors 25:5502. https://doi.org/10.3390/s25175502

[38] Lakshminarayanan B, Pritzel A, Blundell C (2017) Simple and scalable predictive uncertainty estimation using deep ensembles. In: Advances in Neural Information Processing Systems (NeurIPS), https://proceedings.neurips.cc/paper/2017/hash/9ef2ed4b7fd2c810847ffa5fa85bce38-Abstract.html

[39] Lee S, Amakawa S, Yoshida T, et al (2020) 300-GHz CMOS-based wireless link using 40-dBi Cassegrain antenna for IEEE standard 802.15.3d. In: 2020 IEEE International Symposium on Radio-Frequency Integration Technology (RFIT), pp 136–138, https://doi.org/10.1109/RFIT49453.2020.9226231

[40] Lim B, Arik SO, Loeff N, et al (2021) Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting 37:1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

[41] Liu G, He X, Zhao J, et al (2024) Impact of snowfall on terahertz channel performance: Measurement and modeling insights. IEEE Transactions on Terahertz Science and Technology 14:691–698. https://ieeexplore.ieee.org/document/10568390

[42] Liu W, Xia Y, Xu J, et al (2020) Revisiting link quality metrics for wireless sensor networks. In: 2019 IEEE 5th International Conference on Computer and Communications (ICCC), Chengdu, China, https://doi.org/10.1109/ICCC47050.2019.9064098

[43] Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems (NeurIPS), https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[44] Lundberg SM, Erion G, Chen H, et al (2020) From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence 2:56–67. https://doi.org/10.1038/s42256-019-0138-9

[45] Moon EB, Jeon TI, Grischkowsky DR (2015) Long-path THz-TDS atmospheric measurements between buildings. IEEE Transactions on Terahertz Science and Technology 5:742–750. https://ieeexplore.ieee.org/document/7169630

[46] Morshed M, Chowdhury MZ (2026) Curriculum-assisted multi-agent reinforcement learning for scalable V2X resource allocation. Physical Communication 26. https://doi.org/10.1016/j.phycom.2026.103062

[47] Ozpoyraz B, Dogukan AT, Gevez Y, et al (2022) Deep learning-aided 6G wireless networks: A comprehensive survey of revolutionary PHY architectures. IEEE Open Journal of the Communications Society 3:1749–1809. https://doi.org/10.1109/OJCOMS.2022.3210648

[48] Paszke A, Gross S, Massa F, et al (2019) PyTorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems (NeurIPS), https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html

[49] Pedregosa F, Varoquaux G, Gramfort A, et al (2011) Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12:2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html

[50] Petneházi G (2019) Recurrent neural networks for time series forecasting. arXiv:1901.00069

[51] Qiu T, Zhang T, Pan C, et al (2026) Channel measurements and characterization with phase drift compensation for outdoor 330–360 GHz MIMO communications. arXiv:2605.28514

[52] Ravuri S, et al (2021) Skilful precipitation nowcasting using deep generative models of radar. Nature 597:672–677. https://doi.org/10.1038/s41586-021-03854-z

[53] Ribeiro MT, Singh S, Guestrin C (2016) "Why should I trust You?": Explaining the predictions of any classifier. In: Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), pp 1135–1144, https://doi.org/10.1145/2939672.2939778

[54] Saad W, Bennis M, Chen M (2020) A vision of 6G wireless systems: Applications, trends, technologies, and open research problems. IEEE Network 34:134–142. https://doi.org/10.1109/MNET.001.1900287

[55] Serghiou D, Khalily M, Brown TWC, et al (2022) Terahertz channel propagation phenomena, measurement techniques and modeling for 6G wireless communication applications: A survey, open challenges and future research directions. IEEE Communications Surveys & Tutorials 24:1957–1996. https://doi.org/10.1109/COMST.2022.3205505

[56] Shi X, Gao Z, Lausen L, et al (2017) Deep learning for precipitation nowcasting: A benchmark and a new model. In: Advances in Neural Information Processing Systems (NeurIPS), https://proceedings.neurips.cc/paper/2017/hash/a6db4ed04f1621a119799fd3d7545d3d-Abstract.html

[57] Shome S, Das S, Das S, et al (2025) An extensive review of THz communication in 6G: Facilitating technologies with edge computing and native AI. e-Prime – Advances in Electrical Engineering, Electronics and Energy https://www.sciencedirect.com/science/article/pii/S2773186325002221

[58] Srivastava N, Hinton G, Krizhevsky A, et al (2014) Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 15:1929–1958. https://jmlr.org/papers/v15/srivastava14a.html

[59] Stenhammar O, Fodor G, Fischione C (2024) A comparison of neural networks for wireless channel prediction. IEEE Wireless Communications 31:235–241. https://doi.org/10.1109/MWC.006.2300140

[60] Tataria H, Shafi M, Molisch AF, et al (2021) 6G wireless systems: Vision, requirements, challenges, insights, and opportunities. Proceedings of the IEEE 109:1166–1199. https://doi.org/10.1109/JPROC.2021.3061701

[61] Undi F, Schultze A, Keusgen W, et al (2021) Angle-resolved THz channel measurements at 300 GHz in an outdoor environment. In: IEEE International Conference on Communications Workshops (ICC Workshops), https://doi.org/10.1109/ICCWorkshops50388.2021.9473891

[62] Vaswani A, Shazeer N, Parmar N, et al (2017) Attention is all you need. In: Advances in Neural Information Processing Systems (NeurIPS), https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[63] Wang CX, You X, Gao X, et al (2023) On the road to 6G: Visions, requirements, key technologies and testbeds. IEEE Communications Surveys & Tutorials 25:905–974. https://doi.org/10.1109/COMST.2023.3249835

[64] Weng ZK, et al (2021) Millimeter-wave and terahertz fixed wireless link budget evaluation for extreme weather conditions. IEEE Access 9:163476–163491. https://doi.org/10.1109/ACCESS.2021.3132097

[65] Wu H, Li X, Li S, et al (2024) A 260–316 GHz gain-boosted low noise amplifier achieving 17-db gain and 10.5-db noise figure in 130nm SiGe process. In: 2024 IEEE Asia-Pacific Microwave Conference (APMC), pp 1057–1059, https://doi.org/10.1109/APMC60911.2024.10867295

[66] You X, Wang C, Huang J, et al (2021) Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts. Science China Information Sciences 64:110301. https://doi.org/10.1007/s11432-020-2955-6

[67] Zhang Z, Xiao Y, Ma Z, et al (2019) 6G wireless networks: Vision, requirements, architecture, and key technologies. IEEE Vehicular Technology Magazine 14:28–41. https://doi.org/10.1109/MVT.2019.2921208

[68] Zippenfenig P (2023) Open-meteo.com weather API. Version 0.2.69, https://open-meteo.com

Statements and Declarations

Funding. This research did not receive specific grants from funding agencies in the public, commercial or non-profit sectors.

Competing Interests. The authors have no relevant financial or non-financial interests to disclose.

Author Contributions. Israt Jahan: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft. Md. Jahedul Islam: Supervision, Writing – review & editing.

Data Availability. Data will be made available by the corresponding author upon request.

Ethics Approval and Consent to Participate. Not applicable.

Consent for Publication. Not applicable.

Materials Availability. Not applicable.

Code Availability. The code used in this study is available from the corresponding author upon reasonable request.