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.

EpiST-Former: A Meteorological Gated Attention Framework for Multi- Horizon Dengue Outbreak Forecasting and Resource Allocation Support in Bangladesh

Md. Mosharrof Hossain
Mymensingh Engineering College
Md. Rifat Hossen
rifat.220614@s.pust.ac.bd

Pabna University of Science and Technology
Shahreer Islam Sayem
Mymensingh Engineering College
Ozifatun jannat Akhi
Mymensingh Engineering College
Jonayed Al-Faruk
Mymensingh Engineering College

Article

Keywords: Dengue Forecasting, Spatiotemporal Deep Learning, Meteorological Gated Attention, Multi-Horizon Forecasting, Transformer-Based Forecasting, Explainable Artificial Intelligence, Public Health Decision Support

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11653970/v1

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

Additional Declarations: No competing interests reported.

EpiST-Former: A Meteorological Gated Attention Framework for Multi-Horizon Dengue Outbreak Forecasting and Resource Allocation Support in Bangladesh

Md. Mosharrof Hossain[1], Md. Rifat Hossen[2,*], Shahreer Islam Sayem[1],Jonayed Al-Faruk[1], Ozifatun jannat Akhi[1]

[1]*Department of Computer Science and Engineering, Mymensingh Engineering College, Dhaka University, Mymensingh, Bangladesh; mdmosharrofhossain455@gmail.com,ozifa_cse@mec.ac.bd,shahreersayem75@gmail.com*

[2]*Department of Information and Communication Engineering, Pabna University of Science and Technology, Pabna-6600, Bangladesh; rifat.220614@s.pust.ac.bd*

***Corresponding author: Md. Rifat Hossen (rifat.220614@s.pust.ac.bd)**

Abstract

Dengue outbreak forecasting in Bangladesh is challenged by sparse hospital surveillance, temporal variability, and the complex relationships between epidemiological and meteorological factors. This study presents EpiST-Former, a hybrid deep learning framework that combines Bidirectional Gated Recurrent Unit (BiGRU) encoding, a Meteorological Gated Attention (MGA) mechanism, and multi-head self-attention for seven-day-ahead dengue hospital admission forecasting across eleven selected districts of Bangladesh. The framework was developed using Directorate General of Health Services (DGHS) hospital admission records and NASA POWER meteorological data covering 2019–2026. Evaluation was conducted using a chronological test partition spanning 2025–2026. EpiST-Former achieved an MAE of 4.1718, an RMSE of 11.2358, and a raw-scale R^2 of 0.6194. Although these results were slightly better than those of the evaluated rolling persistence baseline in terms of MAE, RMSE, and raw-scale R^2, the persistence baseline achieved a lower sMAPE. Horizon-specific results showed an increase in R^2 from 0.6093 at t+1 to 0.6469 at t+7, although differences in valid observations across horizons require cautious interpretation. Ablation experiments indicated that the log-ratio target formulation, multi-head attention, and meteorological gating were associated with improved performance under the evaluated experimental setting. SHAP analysis identified cumulative rainfall and recent case-history features as important model-attributed predictors. A hospital resource-allocation simulation also suggested a potential reduction in peak bed shortages; however, this result remains simulation-based and requires validation under real-world operational conditions. Overall, EpiST-Former demonstrates potential for short-term dengue hospital admission forecasting, while further validation is needed to assess its robustness and operational applicability.

Keywords: Dengue Forecasting; Spatiotemporal Deep Learning; Meteorological Gated Attention; Multi-Horizon Forecasting; Transformer-Based Forecasting; Explainable Artificial Intelligence; Public Health Decision Support.

1. Introduction

Dengue fever (DF) is a climate sensitive, vector borne viral disease transmitted primarily by *Aedes* mosquitoes [1]. The World Health Organization (WHO) recognizes dengue as one of the ten most significant global public health threats [2], [3], with the disease now present in 125 countries and causing an estimated 400 million infections and 40,000 deaths annually nearly 70% of which occur in endemic tropical and subtropical regions, particularly across Southeast and South Asia [2]. No specific antiviral treatment or universally approved vaccine currently exists for dengue virus (DENV); while the Sanofi Pasteur vaccine has received regulatory approval in 24 countries and has been incorporated into national immunization programs in the Philippines and Brazil, it does not yet offer a comprehensive solution for dengue prevention and control [4]. As a densely populated South Asian country with more than 165 million inhabitants, Bangladesh has experienced recurring dengue outbreaks since 2000, with a marked intensification in outbreak severity and frequency in recent years [5].

In Bangladesh, dengue has become one of the most pressing public health challenges, placing substantial strain on healthcare and surveillance systems[6]. Transmission dynamics in the country are shaped by the interplay of rapid urbanization, climate variability, and evolving viral serotypes, making outbreaks increasingly dynamic, spatially heterogeneous, and difficult to anticipate [7]. Traditional epidemiological surveillance systems often lack the spatiotemporal granularity and predictive

lead time needed to anticipate rapid onset, localized outbreaks, which can delay public health responses, vector control interventions, and healthcare resource preparation during peak epidemic periods.

Over the past decade, a progression of statistical, machine learning, and deep learning approaches has been developed to address these forecasting needs moving from interpretable time-series baselines, to feature-driven ensemble learners, to sequence models capable of learning long-range temporal dependencies, and most recently to attention-based Transformer architectures and graph based spatial models. These developments, together with the growing use of Explainable AI (XAI) techniques to support trust and adoption in public health settings, are reviewed in detail in Section 2 (Related Work).

Despite this progress, four gaps remain across the existing literature:

1. **Dynamic meteorological feature fusion:** most models incorporate meteorological variables through direct concatenation or fixed fusion strategies, even though the influence of climate variables plausibly varies across transmission phases and seasonal conditions.
2. **Epidemiologically sensitive loss optimization:** conventional objectives such as Mean Squared Error (MSE) treat all prediction errors uniformly, whereas underestimating outbreak peaks carries disproportionate public-health cost.
3. **Dynamic spatial dependency modeling:** many frameworks analyze districts in isolation or rely on predefined spatial relationships, limiting their ability to capture evolving inter-regional transmission patterns.
4. **The predictive-to-prescriptive decision gap:** relatively few studies translate outbreak forecasts into actionable, resource-planning support for public-health decision-makers.

To address these gaps, this study introduces EpiST-Former, an end-to-end forecasting framework that combines recurrent temporal encoding, an early Meteorological Gated Attention (MGA) mechanism, multi-head self-attention, and a reinforcement-learning-informed decision-support module. The framework integrates adaptive meteorological–epidemiological feature fusion, an epidemiologically sensitive loss formulation, multi-horizon (7-day) forecasting, SHAP-based interpretation, and a downstream hospital resource-allocation simulation into a single pipeline.

The main contributions of this study are summarized as follows:

1. Meteorological Gated Attention (MGA) a dual-stream fusion mechanism that adaptively integrates meteorological information with historical dengue incidence patterns, with a learned gate regulating the contribution of meteorological signals within the fused representation.
2. **Long-range temporal self-attention modeling:** multi-head self-attention applied to the fused representations to capture long-range temporal dependencies across the input sequence.
3. **Epidemiologically sensitive loss optimization:** a masked Huber log-ratio loss that gives greater weight to critical outbreak deviations and incorporates constraints relevant to epidemic forecasting.
4. **Downstream resource-allocation support:** use of the model's forecasting outputs in a hospital resource-allocation simulation to examine their potential value for public-health planning.

The remainder of this paper is organized as follows: Section 2 reviews related work across statistical, machine learning, deep learning, Transformer-based, spatial, physics-informed, and explainable forecasting approaches; Section 3 describes the data pipeline and the proposed EpiST-Former architecture; Section 4 presents the experimental results; and Section 5 discusses the findings, practical implications, and study limitations.

2. Related Work

2.1 Statistical Forecasting Approaches

Statistical models represent some of the earliest dengue forecasting frameworks, valued for their simplicity, interpretability, and computational efficiency. Classical methods, including Autoregressive Integrated Moving Average (ARIMA) and its seasonal extensions SARIMA and SARIMAX, have been widely used as baseline forecasting tools, primarily analyzing historical case patterns, temporal trends, and seasonal variation, with SARIMAX additionally incorporating external climate variables. In Bangladesh, Khondaker and Kamrujjaman [8] evaluated SARIMAX and Poisson regression models and found that incorporating climate-lag effects, such as rainfall and temperature variation, improved forecasting stability. Sharmin Akther and Md. Al-Mamun compared SARIMA with several machine-learning models for dengue forecasting in Dhaka and reported that Random Forest provided the strongest predictive performance under chronological validation [REF-NEEDED]. Simple statistical benchmarks such as Moving Average and Naive Seasonal Baseline are also commonly used to gauge the added value of more advanced models. Statistical approaches offer high interpretability and reliable baseline performance, and their relatively limited capacity to represent nonlinear, abrupt epidemic transitions has motivated the development of machine-learning-based forecasting approaches.

2.2 Machine Learning Forecasting Approaches

Machine learning methods extend statistical baselines by capturing nonlinear relationships between climatic factors and dengue incidence. Ensemble decision-tree models such as Random Forest are effective at capturing nonlinear interactions and estimating feature importance, as demonstrated by Dewi et al. [9]. Support Vector Regression (SVR) uses kernel-based functions to model nonlinear relationships and has shown effectiveness on small- to medium-scale epidemiological datasets, though it faces scalability constraints on large, high-dimensional spatiotemporal problems. More recently, gradient-boosting frameworks such as XGBoost and LightGBM have shown strong performance in epidemic forecasting: XGBoost has been applied to dengue forecasting by Rahman and Shiddik [7] and Dewi et al. [9], while LightGBM was used by Rahman and Shiddik [7] to identify important outbreak drivers in Bangladesh. These models typically rely on manually engineered temporal features and do not inherently learn long-range dependencies from sequential data, motivating the adoption of deep-learning-based temporal forecasting approaches.

2.3 Deep Sequence Forecasting Approaches

Deep learning architectures have been widely adopted for dengue forecasting because of their ability to learn complex temporal dependencies directly from historical incidence and meteorological sequences [10]. Long Short-Term Memory (LSTM) networks are extensively used for their gated memory mechanism; in the Bangladesh context, Mobin and Begum [11] showed that recurrent architectures including LSTM, GRU, BiLSTM, and BiGRU effectively capture temporal relationships between meteorological variables and dengue incidence. Gated Recurrent Units (GRUs) offer a computationally efficient alternative to LSTMs with fewer parameters while maintaining competitive performance [10], [11]. Building on these, bidirectional models such as BiLSTM with Attention leverage both forward and backward temporal context while emphasizing informative historical time steps [11], [12], and Temporal Convolutional Networks (1D-TCNs) use causal, dilated convolutions to capture long-range patterns with parallelizable computation [12]. Hybrid architectures combining convolutional and recurrent layers have also been proposed to jointly capture local fluctuations and long-term epidemic trends [13]. A remaining consideration for these architectures is that sequential hidden-state updates constrain computational parallelism for long forecasting horizons [10], [12] and meteorological variables are typically incorporated as static input features rather than being dynamically weighted across different transmission phases [11], [12]

2.4 Transformer-Based and Spatially-Aware Forecasting

Transformer-based architectures have emerged as promising alternatives for epidemic time-series forecasting, as self-attention allows greater parallelization than recurrent processing and directly models relationships among distant elements of an input sequence [14]. The original Transformer introduced multi-head self-attention and positional encoding, reducing sequential dependency while enabling flexible long-range modeling [14]; a known characteristic of vanilla self-attention is its quadratic computational and memory complexity with respect to sequence length [14], [15]. In the dengue context, Nguyen et al. [16] evaluated multiple deep-learning architectures using climate data from 20 provinces in Vietnam, providing dengue-specific evidence for comparing attention-based and recurrent architectures under climate-informed settings.

To represent transmission across geographically connected regions, recent work has extended deep-learning forecasting with spatial graphs and dynamic disease mechanisms. Wang and Jin [17] proposed a multi-region infectious-disease forecasting framework combining spatiotemporal graph learning with metapopulation disease dynamics, and Jiao et al. [18] investigated mobility-informed spatiotemporal epidemic forecasting using attention-based sequence models, both highlighting the value of integrating spatial connectivity with temporal disease signals. Cross-attention mechanisms provide an established approach for integrating heterogeneous data sources used by Prabu et al. [19] to model lagged climate–disease relationships and by Yu et al. [20] to integrate spatiotemporal representations though this approach requires separate projections for each input stream, adding computational overhead. Building on this line of work, EpiST-Former instead employs an early **Meteorological Gated Attention (MGA)** mechanism that adaptively integrates epidemiological and meteorological representations into a unified latent space *before* multi-head self-attention is applied, allowing the model to dynamically regulate the contribution of meteorological information while capturing long-range temporal dependencies from a shared representation.

Physics-informed learning has also been explored to improve the mechanistic consistency of epidemic forecasting: Shaier and Raissi [21] introduced Disease-Informed Neural Networks that incorporate disease-transmission dynamics into neural-network learning through physics-based constraints, providing a foundation for integrating data-driven learning with epidemiological knowledge. Recent dengue-focused research has further examined physics-informed modeling of unobserved vector dynamics, though such findings are best interpreted as preprint evidence pending peer-reviewed validation [22].

Overall, Transformer-based models generally benefit from careful regularization and sufficient training data to achieve stable representations in limited regional epidemiological datasets [15], and standard architectures do not inherently encode geographic relationships, seasonal transmission mechanisms, or biological disease constraints unless explicitly augmented with graph-based, mobility-aware, or physics-informed components [17], [21]. These observations motivate integrated forecasting frameworks such as the one proposed in this study that combine long-range temporal modeling with adaptive meteorological integration.

2.5 Explainable AI in Epidemic Forecasting

Beyond predictive accuracy, model interpretability is important for practical adoption in public-health decision-making, since stakeholders need to understand, evaluate, and appropriately trust model outputs [23], [24]. While advanced deep-learning and Transformer-based architectures can capture complex nonlinear relationships in epidemic dynamics, their relatively opaque internal representations can make transparent interpretation more difficult, motivating the growing use of Explainable Artificial Intelligence (XAI) techniques [23], [24]. Among these, SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) have been widely adopted for interpreting machine-learning and deep-learning models through feature attribution and local surrogate explanations, respectively [25], [26].

Recent dengue studies illustrate the value of XAI approaches for translating complex predictive outputs into interpretable epidemiological insight. Rahman and Shiddik [7] applied SHAP and LIME to examine the contribution of climatic, socio-demographic, and landscape-related variables to dengue outbreak prediction in Bangladesh, identifying population density, relative humidity, and minimum temperature as important model-attributed predictors. District-level dengue early-warning research has further integrated explainable AI with Bayesian deep learning to support interpretable risk assessment and local preparedness planning [27], and complementary work has discussed the importance of responsible interpretation and transparency when applying XAI in dengue forecasting systems [7]. It is worth noting that SHAP and LIME are model-explanation techniques rather than direct evidence of causal relationships feature attribution describes how a model uses available information for prediction but does not, by itself, establish biological causation [26], [27]. In this study, SHAP-based interpretation is therefore used to complement, rather than replace, epidemiological knowledge, and is applied to EpiST-Former to characterize the relative contribution of meteorological, biological, and static demographic feature streams to the model's forecasts.

2.6 Positioning of This Study

Taken together, the reviewed literature reflects steady progress from interpretable statistical baselines, through feature-engineered machine learning models, toward deep sequence and attention-based architectures capable of learning long-range temporal structure, alongside a growing emphasis on model interpretability and, more recently, spatial and physics-informed extensions. Building on this progression, EpiST-Former combines four complementary elements within a single framework an early Meteorological Gated Attention (MGA) mechanism, multi-head self-attention over the fused representation, an epidemiologically sensitive Huber log-ratio loss, and SHAP-based interpretation linked to a downstream resource-allocation simulation with the goal of jointly addressing adaptive meteorological integration, epidemic-risk-aware optimization, and translational decision support within a single dengue forecasting pipeline for Bangladesh. Several previous studies have investigated dengue forecasting using statistical, machine-learning, and deep-learning approaches. A summary of these studies, including their key contributions, limitations, and relevance to the proposed study, is presented in Table 1.

Table 1. Literature Review of Dengue Forecasting, Deep Learning, and Explainable AI Approaches

Author [Ref.]	Study / Approach	Key Contribution / Finding	Limitation / Gap	Relevance to Our Study
Khondaker & Kamrujjaman [8]	SARIMAX, Poisson regression; climate-lag features	Climate-lag effects such as rainfall and temperature variation improved forecasting stability	Limited ability to model nonlinear epidemic transitions	Motivates climate-aware deep-learning forecasting
Dewi et al. [9]	Machine-learning models	Applied ML approaches for predicting dengue cases and severity	Feature-driven ML does not inherently capture long-range sequential dependencies	Provides ML baseline/reference
Lober et al. [10]	Systematic review of neural networks	Reviewed neural-network approaches for dengue prediction	Highlights limitations and variability across existing approaches	Supports deep sequence modeling
Mobin & Begum [11]	LSTM, GRU, BiLSTM, BiGRU	Captured relationships between meteorological variables and dengue incidence	Sequential processing limits parallelism; meteorological variables are generally treated as input features	Motivates adaptive meteorological fusion and attention
Fan et al. [12]	Spatio-temporal attention-based TCN	Combines temporal convolution and attention for multivariate time-series prediction	Does not specifically address adaptive meteorological gating for dengue	Supports temporal attention-based modeling
Zhao et al. [13]	Hybrid deep-learning architecture	Captures local fluctuations and longer-term epidemic patterns	Hybrid sequence models remain computationally constrained	Supports hybrid temporal representation learning
Vaswani et al.	Transformer / Multi-head self-	Enables long-range	Standard self-	Foundation for

Author [Ref.]	Study / Approach	Key Contribution / Finding	Limitation / Gap	Relevance to Our Study
[14]	attention	sequence modeling through self-attention	attention has quadratic complexity	Transformer component
Nguyen et al. [16]	Climate-informed deep learning	Evaluated deep-learning models for dengue forecasting using climate data	Different geographic and epidemiological setting	Supports climate-informed dengue forecasting
Wang & Jin [17]	Spatiotemporal graph learning	Integrates spatial graph learning with disease dynamics	Requires explicit spatial/dynamic modeling	Motivates spatially aware forecasting
Jiao et al. [18]	Mobility-informed LSTM + attention	Models mobility-informed spatiotemporal epidemic patterns	Relies on mobility information	Supports spatial-temporal epidemic modeling
Prabu et al. [19]	Domain-knowledge-informed Transformer	Models lagged climate–disease relationships	Cross-attention requires separate projections for input streams	Closely related to climate–disease fusion
Yu et al. [20]	Cross-attention fusion + spatial-temporal GCN	Integrates heterogeneous spatiotemporal representations	Separate projections increase architectural complexity	Supports heterogeneous feature fusion
Shaier & Raissi [21]	Disease-Informed Neural Networks	Incorporates disease-transmission dynamics into neural learning	Requires mechanistic constraints/assumptions	Supports epidemiologically informed learning
Cao et al.[22]	Physics-informed neural networks	Models unobserved vector dynamics for dengue forecasting	Evidence is currently preprint-based	Shows emerging physics-informed direction
Zhou et al. [15]	Informer / efficient Transformer	Efficient long-sequence time-series forecasting	Requires suitable regularization and sufficient data	Supports efficient temporal modeling
Holzinger et al. [23]	Explainability in medical AI	Discusses explainability and causability	Explainability does not establish causality	Foundation for responsible XAI
Amann et al. [24]	Explainable AI in healthcare	Provides multidisciplinary perspective on healthcare AI explainability	Challenges remain in trustworthy interpretation	Supports XAI component

Author [Ref.]	Study / Approach	Key Contribution / Finding	Limitation / Gap	Relevance to Our Study
Lundberg & Lee [25]	SHAP	Provides feature-attribution framework for model interpretation	Attribution does not establish causality	Basis for SHAP interpretation
Ribeiro et al. [26]	LIME	Provides local surrogate explanations	Local explanations may not represent global model behavior	Complementary XAI methodology
Rahman & Shiddik [7]	XAI for dengue forecasting	Applies explainable AI to dengue prediction	Interpretation remains distinct from causal inference	Supports responsible XAI

3. Methodology

3.1 Overview of the preprocessing pipeline

The pipeline transforms heterogeneous dengue surveillance records into a district-level daily panel and subsequently into temporally ordered model-ready tensors. The final analytical scope comprises 11 focal districts Barishal, Chattogram, Dhaka, Faridpur, Gazipur, Gopalganj, Khulna, Mymensingh, Rajshahi, Rangpur, and Sylhet covering 27 August 2019 to 7 July 2026. The complete district-day grid therefore contains 27,577 structural slots (11 districts × 2,507 days).Hospital surveillance data were obtained from the Directorate General of Health Services (DGHS), Bangladesh [28]. The complete dataset preparation workflow, including data extraction, harmonization, missingness handling, feature engineering, sequence construction, and chronological dataset partitioning, is illustrated in Fig. 1. The complete step-by-step procedure followed for data preparation, validation, feature engineering, scaling, chronological splitting, sequence generation, and final dataset locking is summarized in Table 2.

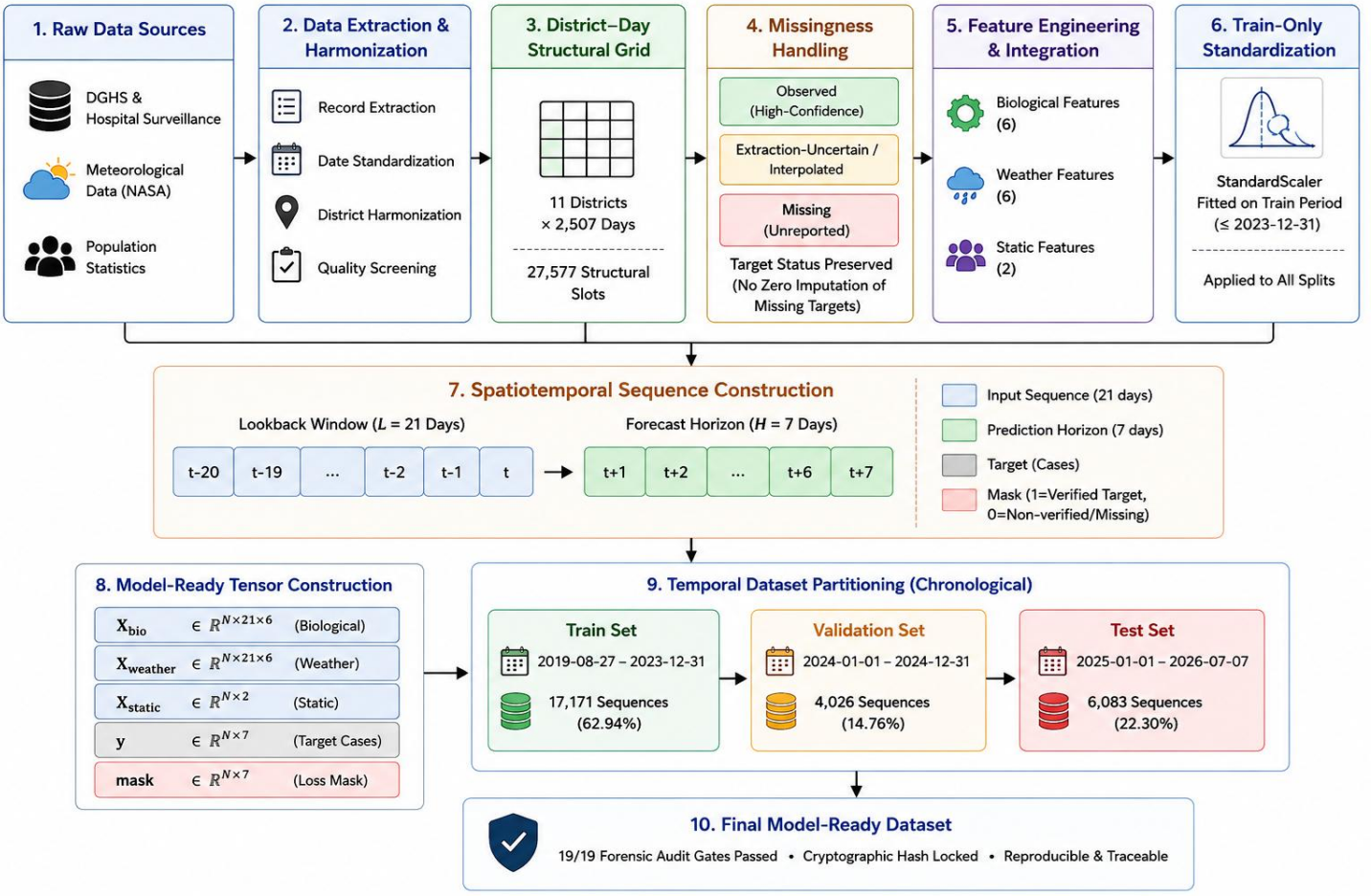

Fig. 1. Dataset preparation final pipeline.

Table 2. Step by step procedure of data preparation

Stage	Operation	Primary output
1	Raw surveillance extraction and provenance preservation	dghs_extracted_raw.csv
2	Target/master construction and district-day grid generation	dengue_surveillance_grid_v1.csv
3	Structural and missingness validation	validation_report_v1.txt
4	Feature engineering and independent forensic recomputation	Audit evidence files
5	Scaling, chronological splitting, and sequence generation	model_ready_tensors.npz (finalized as final_model_ready_v1.npz after locking)
6	Final lock and reproducibility audit	FINAL_DATASET_LOCK_v1.md and audit outputs

3.1.1 Raw Surveillance Extraction and Provenance Preservation

Raw epidemiological observations were extracted without interpolating or zero-filling missing reports. The extraction script explicitly preserved the raw district name, observation/publication date, original case value, aggregation level, extraction method, confidence label, source reference, and observation status.

For Bijoy-decoded surveillance records, the script reads bijoy_master_raw.csv, strips the raw district and date fields, converts the case field to numeric values where possible, and assigns records to either a district-level category or an OUTSIDE_DHAKA_CITY aggregate category. Numeric observations are labelled OBSERVED at this source-record level, whereas unparseable or absent values are retained as NaN and labelled EXTRACTION_UNCERTAIN; this record-level label is distinct from the grid-level data_status field introduced in Section 3.1.2. Source-level provenance is retained through a source identifier and row-level source reference.

Hospital surveillance data are read using Python's CSV reader to preserve quoted comma-containing fields. The date, facility name, district, and case count are extracted, with the district used as the geographic identifier when available. These observations are labelled at the HOSPITAL aggregation level and receive HIGH or UNCERTAIN extraction confidence depending on successful numeric conversion.

Importantly, this stage does not convert unreported observations into zero cases. This distinction is essential because a zero is an epidemiological value, whereas a missing report represents an absence of observation.

3.1.2 Construction of the District-Day Surveillance Grid

The target-processing stage preserves three analytically distinct states: verified observations, missing observations, and records already labelled as interpolated in the source dataset. When a source record is marked as observed, the case value is retained and classified as VERIFIED_NONZERO or VERIFIED_ZERO. Records already identified as interpolated retain their associated interpolation metadata and remain analytically distinguishable from verified observations and missing records.

A complete daily index was then generated through the Cartesian product of the study dates and focal districts. The grid spans 27 August 2019 to 7 July 2026 and contains one row for every district-day combination. When a district-day was absent from the target source, reported_cases was assigned NaN, data_status was set to MISSING, and source metadata were set to NONE. Verified observations were assigned data_status OBSERVED with HIGH confidence, whereas records carrying interpolation-related source metadata were retained separately as data_status EXTRACTION_UNCERTAIN with MEDIUM confidence in the surveillance grid. This grid-level data_status field operates alongside, and should not be conflated with, the case-level VERIFIED_NONZERO/VERIFIED_ZERO classification described above and the source-record label from Section 3.1.1.

District-level missingness metrics were subsequently calculated from the structural grid. These included total slots, observed slots, uncertain/interpolation-related slots, missing slots, missing percentage, longest missing streak, median missing streak, and counts of one-, two-, three-, and greater-than-three-day gaps..

3.1.3 Structural validation

The generated surveillance grid was validated before feature construction. The validation procedure checked: (i) the exact row count of 27,577; (ii) the presence of exactly 11 focal districts; (iii) the date bounds of 2019-08-27 to 2026-07-07; (iv) the invariant that MISSING rows contain NaN reported_cases; and (v) the absence of negative case counts.

The supplied validation report records PASS for all five checks. The audited status distribution was 4,970 OBSERVED slots (18.02%), 306 EXTRACTION_UNCERTAIN slots (1.11%), and 22,301 MISSING slots (80.87%).

3.1.4 Environmental and static data integration

For model preparation, the target dataset was merged with district-level weather variables obtained from the NASA POWER daily API [29] using district and date as the joint key. The merged weather variables were Temperature_C, Humidity_Percent, Rainfall_mm, Temp_roll_7d, Humidity_roll_7d and Rainfall_roll_14d. District population was mapped from a population table[30], transformed using log1p(Population), and stored as Log_Population. Each study district was also assigned a deterministic integer District_ID according to the sorted district list.

3.1.5 Biological and weather feature engineering

Feature engineering was performed independently within each district after chronological sorting, thereby preventing lag and rolling-window calculations from crossing district boundaries.

Table 3. Feature engineering outline.

Feature group	Features	Construction
Biological lags	Cases_lag_7, Cases_lag_14, Cases_lag_21	District-wise shift of final_target_value by 7, 14, and 21 days
Biological rolling	Cases_roll_mean_7d	Trailing 7-day mean with min_periods=1
Biological rolling	Cases_roll_std_7d	Trailing 7-day sample standard deviation; initial undefined value filled with 0
Reporting history	days_since_last_report	Counter reset on VERIFIED_NONZERO or VERIFIED_ZERO; incremented otherwise
Weather rolling	Temp_roll_7d	Trailing 7-day mean
Weather rolling	Humidity_roll_7d	Trailing 7-day mean
Weather rolling	Rainfall_roll_14d	Trailing 14-day sum
Static	Log_Population, District_ID	District-level time-invariant covariates

After feature construction, missing biological lag/rolling values and missing weather feature values were filled with 0.0 in the model-input table. According to the supplied audit logic, lag-related zero filling at the beginning of each district series is treated as structural initialization rather than as replacement of unreported target observations. The reporting target itself remains separately represented by its status and mask.

3.1.6 Feature scaling

Biological and weather features were standardized using separate sklearn StandardScaler objects. Both scalers were fitted only on rows dated on or before 31 December 2023. The fitted scalers were then applied to all chronological splits. Static features were not included in these two scaler groups.

This train-only fitting strategy prevents validation and test feature distributions from contributing to the estimated training mean and variance.

3.1.7 Chronological sequence generation

The model-ready representation uses a lookback window L=21 days and a forecast horizon H=7 days. For each district, each sample contains a 21-day input window ending at time t and a seven-day target vector covering t+1 through t+7. Separate input tensors are created for biological features, weather features, and static features.

Samples are assigned chronologically according to the end date of the forecast horizon: training if the final target date is on or before 2023-12-31, validation if it is on or before 2024-12-31 but later than the training boundary, and testing otherwise.

A target mask is generated for every forecasted day. VERIFIED_NONZERO and VERIFIED_ZERO targets receive mask value 1.0; all other target statuses receive mask value 0.0. Thus, unobserved or non-verified target positions remain in the tensor but can be excluded from loss contribution.

3.1.8 Model-ready tensor structure

The final archive contains 15 arrays: X_bio, X_weather, X_static, y, and mask for each of the train, validation, and test partitions. The biological and weather input tensors have shape (N, 21, 6), static inputs have shape (N, 2), and target and mask tensors have shape (N, 7).

Table 4. Data set splitting ratio for training model

Split	Sequences	X_bio / X_weather	X_static	y / mask
Train	17171	(N, 21, 6)	(N, 2)	(N, 7)
Validation	4026	(N, 21, 6)	(N, 2)	(N, 7)
Test	6017	(N, 21, 6)	(N, 2)	(N, 7)

3.1.9 Independent forensic verification

The feature-engineering audit independently recomputes the 7-, 14-, and 21-day case lags; weather rolling statistics; biological rolling statistics; scaler fitting population; calendar continuity; masked-loss gradient behavior; tensor values; and feature ordering. The audit also checks target hashing, district scope, temporal continuity, split boundaries, and reproducibility-related properties.

The supplied loss audit defines a masked weighted error in which sample_weight multiplies the loss weights. A unit test uses automatic differentiation when TensorFlow is available and verifies zero gradient contribution for masked target positions. This audited formulation is a representative masked weighted squared-error function used to validate the masking mechanism in principle; it is architecturally distinct from the masked Huber log-ratio loss that trains the final EpiST-Former model (Section 3.2.6). Accordingly, this audit certifies that masked positions do not influence the tested gradient under the audited formulation, rather than certifying gradient behavior for the specific loss instance used in final model training.

Tensor reconciliation compares the first biological training sequence against the corresponding scaled dataframe window and records the maximum absolute difference. Feature ordering is separately enumerated for all 14 model inputs: six biological, six weather, and two static features.

3.1.10 Reproducibility and dataset locking

The final lock audit checks dataset lineage, district scope, master-input row count, lag availability, rolling-feature gates, scaler training scope, calendar continuity, loss-mask behavior, tensor integrity, feature mapping, and cryptographic signatures. SHA-256 hashes are generated for the target dataset, master input CSV, and model-ready tensor archive.

The supplied lock certificate identifies the frozen target dataset hash as

ef309fc618afc7e76f3190546955ecf5f4825f11ee505cc1ec6c1d0e511a4c6f and specifies L=21 and H=7.

3.2 Proposed EpiST-Former model architecture and training protocol

Building on the finalized tensor archive described in Section 3.1.8 (final_model_ready_v1.npz), the training script implements the proposed EpiST-Former forecasting model together with its data loading, log-ratio target transformation, masked loss optimization, evaluation, and figure-generation procedures. The model consumes the biological, weather, and static input tensors and their associated masks (Section 3.1.8, Table 4) and produces seven-day-ahead forecasts corresponding to the horizon H = 7 defined in Section 3.1.7. The overall architecture of the proposed EpiST-Former model, including the biological, weather, and static input branches, Meteorological Gated Attention (MGA), multi-head self-attention, feature fusion, and the final 7-day forecasting head, is illustrated in Fig. 2.

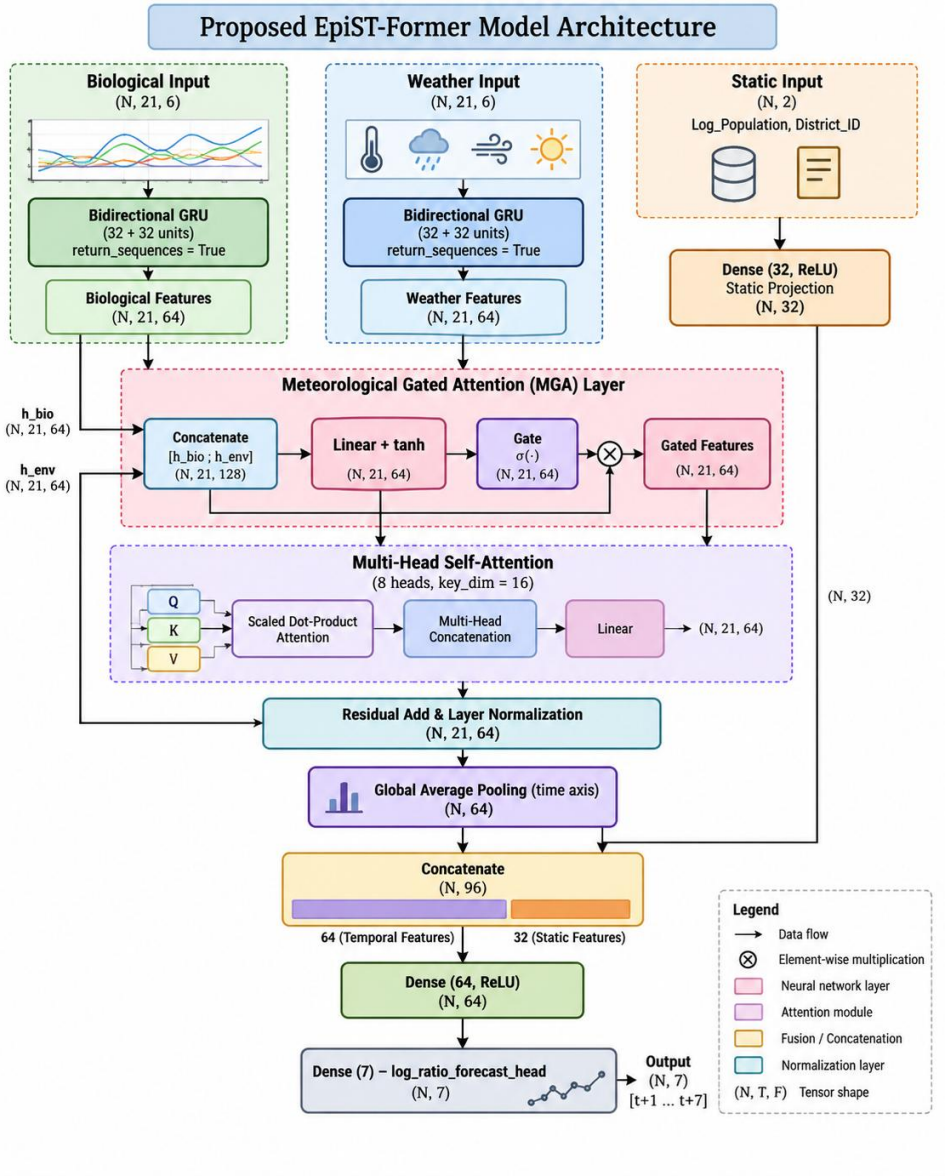

Fig. 2. Proposed EpiST-Former model architecture.

3.2.1 Dataset discovery and reproducibility configuration

Global random seeds for NumPy and TensorFlow are fixed at the beginning of the script (seed = 42) prior to any model construction, consistent with the reproducibility orientation established during dataset locking in Section 3.1.10. A dedicated discovery routine searches a prioritized list of candidate paths a recursive Kaggle input search, the current working directory, the parent directory, and an explicit local pipeline directory for the finalized archive final_model_ready_v1.npz introduced in Section 3.1.8, and raises a FileNotFoundError if the archive cannot be resolved from any candidate location. The first matching path is reported to the console before loading proceeds.

3.2.2 Input tensors and log-ratio target construction

The train, validation, and test partitions of X_bio, X_weather, X_static, y, and the target mask are loaded directly from the archive, matching the shapes reported in Table 4. Rather than forecasting raw case counts directly, the model is trained on a log-ratio target defined relative to a per-sequence reference level. This reference level, termed the log base, is derived from the trailing seven-day mean case count (Cases_roll_mean_7d, the fourth of the six biological features enumerated in Table 3, Section 3.1.5) at the final timestep of each 21-day input window. Because this feature was standardized during Section 3.1.6, it is first inverse-transformed using the fixed scaler statistics recorded in the script (mean = $5.01324872 \times 10^{-4}$, standard deviation = $1.29299186 \times 10^{-2}$) and clipped at zero. The log base is then computed as log1p of this recovered

rolling mean, and the training and validation targets are expressed as log1p(max(y, 0)) minus the log base, yielding a log-ratio that measures the proportional departure of each forecasted day from the sequence's recent case level. Static covariates are standardized separately using the mean and standard deviation of the training-fold static features only, applied identically to the validation and test partitions, mirroring the train-only fitting strategy applied to the biological and weather scalers in Section 3.1.6. For model fitting, the log-ratio target and the per-day mask are concatenated along the feature axis into a single (N, 14) array, allowing the mask to travel through the Keras training loop alongside the target it qualifies.

3.2.3 Meteorological Gated Attention (MGA) layer

The biological and weather input sequences are each processed by an independent Bidirectional GRU (32 units per direction, return_sequences = True), producing 64-dimensional per-timestep representations across the 21-day lookback window. These two sequences are then combined by a custom Meteorological Gated Attention layer, which computes a sigmoid gate from the weather representation and separate tanh projections of the biological and weather representations, and returns their gate-weighted convex combination at every timestep. A gate value close to one indicates that the fused representation is weighted more heavily toward meteorological information at that timestep, whereas a value close to zero up-weights the biological and epidemiological history channel. The layer preserves the temporal dimension, producing a gated sequence of shape (N, 21, 64).

3.2.4 Multi-head self-attention, residual connection, and pooling

The gated sequence is passed through a multi-head self-attention block (8 heads, key dimension 16) in which the query, key, and value tensors are all set to the gated features, allowing the model to weight informative days within the 21-day window relative to one another. The attention output is added to the gated features through a residual connection and normalized with layer normalization. Global average pooling over the time axis then compresses the attended 21-timestep sequence into a single 64-dimensional representation per sample.

3.2.5 Static covariate fusion and prediction head

The standardized static covariates (Log_Population and District_ID, Section 3.1.4) are projected through a Dense(32, ReLU) layer and concatenated with the pooled 64-dimensional temporal representation, producing a 96-dimensional joint representation. This is passed through a Dense(64, ReLU) layer and a final linear Dense(7) output layer, named log_ratio_forecast_head, which produces one log-ratio value for each of the seven forecast horizon days (t+1 through t+7), consistent with the forecast horizon H = 7 defined in Section 3.1.7.

3.2.6 Masked Huber log-ratio loss function

The model is optimized using a custom masked Huber loss defined over the log-ratio target. At each call, the combined (N, 14) label tensor is split into the first seven columns, containing the true log-ratio values, and the last seven columns, containing the corresponding per-day mask. An elementwise Huber loss (with a configurable delta parameter, default 1.0) is computed between the true and predicted log-ratio values, multiplied by the mask, and normalized by the sum of mask entries plus a small epsilon term rather than by the total number of elements. Consequently, positions that are not VERIFIED_NONZERO or VERIFIED_ZERO under the masking convention introduced in Section 3.1.7 make no contribution to the training gradient, extending the masking principle established during data preparation into the loss computation itself. This masking mechanism is structurally analogous to but implements a distinct error formulation from the masked weighted-error function independently verified in the forensic audit (Section 3.1.9); the zero-gradient property demonstrated there applies to the masking principle in general and has not been separately re-verified against this specific Huber log-ratio loss instance.

3.2.7 Training configuration

The model is compiled with the Adam optimizer (initial learning rate 0.001) and the masked Huber log-ratio loss, and trained for up to 40 epochs with a batch size of 32 on the 17,171 training sequences reported in Table 4, using the validation

partition for monitoring. Three callbacks govern the training loop: a model checkpoint that retains only the weights achieving the lowest validation loss; early stopping with a patience of 12 epochs, which restores the best observed weights upon termination; and a learning-rate reduction callback that halves the learning rate after five epochs without validation-loss improvement, subject to a floor of 1×10^{-5}. The best checkpoint is written to publication_artifacts/best_epist_former_model.keras. The training hyperparameters and callback configurations used for optimizing the proposed EpiST-Former model are summarized in Table 5.

Table 5. Training hyperparameters and callback configuration

Hyperparameter / Callback	Configuration
Optimizer	Adam, learning rate = 0.001
Loss function	Masked Huber log-ratio loss ($\delta = 1.0$)
Epochs / batch size	40 / 32
Model checkpoint	monitor = val_loss, save_best_only = True
Early stopping	monitor = val_loss, patience = 12, restore_best_weights = True
Learning-rate reduction	monitor = val_loss, factor = 0.5, patience = 5, min_lr = 1e-5

3.2.8 Inference, inverse transformation, and evaluation protocol

At inference, the trained model predicts log-ratio values for the test partition, which are converted back to case-count scale by adding the test-set log base derived in Section 3.2.2, clipping at zero, and applying the inverse log1p transform (expm1). Global evaluation metrics mean absolute error, root mean squared error, symmetric mean absolute percentage error, and coefficient of determination (R^2) computed on both raw and log1p-transformed case counts are calculated only over positions where the test mask exceeds 0.5, restricting evaluation to VERIFIED_NONZERO and VERIFIED_ZERO target days. A parallel per-horizon evaluation repeats the mean absolute error, root mean squared error, and R^2 calculations independently for each of the seven forecast days (t+1 through t+7) under the corresponding per-day mask, and the resulting table is written to epist_former_per_horizon_metrics.csv within a publication_artifacts output directory.

3.2.9 Diagnostic and publication figures

In addition to the numerical outputs, the script generates four diagnostic figures at 300 dpi using a consistent seaborn visual theme: a training and validation loss curve across epochs; an R^2 decay curve summarizing forecast performance across the seven forecast horizons; a scatter plot of predicted against actual test-set case counts with a 1:1 reference line; and an overlay time series comparing observed cases against the seven-day-ahead forecast for the first 300 test sequence points. These figures, together with the per-horizon metrics table, are stored in the publication_artifacts output directory as part of the model's reporting output.

4. Results

4.1 Comparative Benchmark Performance on Prospective Epidemic Test Data

Table 6 summarizes the performance of EpiST-Former and the evaluated baseline models on the chronological test partition. EpiST-Former achieved an MAE of 4.1718, an RMSE of 11.2358, a raw-scale R^2 of 0.6194, and a log-scale R^2 of 0.5817. Among the evaluated models, it obtained the lowest MAE and RMSE and the highest raw-scale R^2. However, the improvement over the 7-day rolling persistence baseline was relatively small: the persistence model achieved an MAE of 4.2447, an RMSE of 11.2850, and an R^2 of 0.6160. In addition, the persistence baseline produced a lower sMAPE (75.14%) than EpiST-Former (118.08%). These results indicate that EpiST-Former provides modest improvements in absolute error

and raw-scale explained variance over the persistence baseline, while its relative percentage error remains comparatively high. The model also outperformed the Random Forest and Ridge regression models on the reported MAE, RMSE, and raw-scale R^2 metrics. Therefore, its performance advantage should be interpreted in relation to the specific evaluation metrics rather than as uniform superiority across all criteria. The relatively high sMAPE values, particularly for EpiST-Former, should be interpreted in the context of the sparse and zero-heavy surveillance observations. Percentage-based error measures can become unstable when actual case counts are zero or close to zero. Consequently, sMAPE may provide a less favorable assessment than absolute error measures in this dataset. To improve transparency, the number of valid observations used for each metric and the exact sMAPE implementation should be reported. MAE and RMSE are therefore also presented to describe prediction errors in the original hospital-admission count scale.

Table 6. Master benchmark comparison matrix evaluated on the prospective test set (6,017 samples, 11 districts, 2025–2026).

Model Name	**Model Family**	**Test MAE ↓**	**Test RMSE ↓**	**sMAPE (%) ↓**	**R^2 Raw ↑**	**R^2 Log ↑**	**t+1 R^2 ↑**	**t+7 R^2 ↑**
Naive Zero Baseline	Statistical Baseline	6.6922	19.4025	66.10%	-0.1350	-0.3723	-0.1342	-0.1361
Historical Train Mean	Statistical Baseline	6.7014	19.3932	199.25%	-0.1339	-0.3476	-0.1331	-0.1350
7-Day Rolling Persistence	Statistical Baseline	4.2447	11.2850	75.14%	0.6160	0.5421	0.6057	0.6413
Ridge Regressor (Scale-Adaptive)	Classical ML	153.4708	1426.7078	119.42%	-6136.0610	-0.7735	-4595.8774	-15939.4824
Random Forest Regressor	Classical ML	4.8038	12.6087	144.72%	0.5207	0.4868	0.4987	0.6168
Proposed EpiST-Former	**Proposed Framework**	4.1718	11.2358	118.08%	0.6194	0.5817	0.6093	0.6469

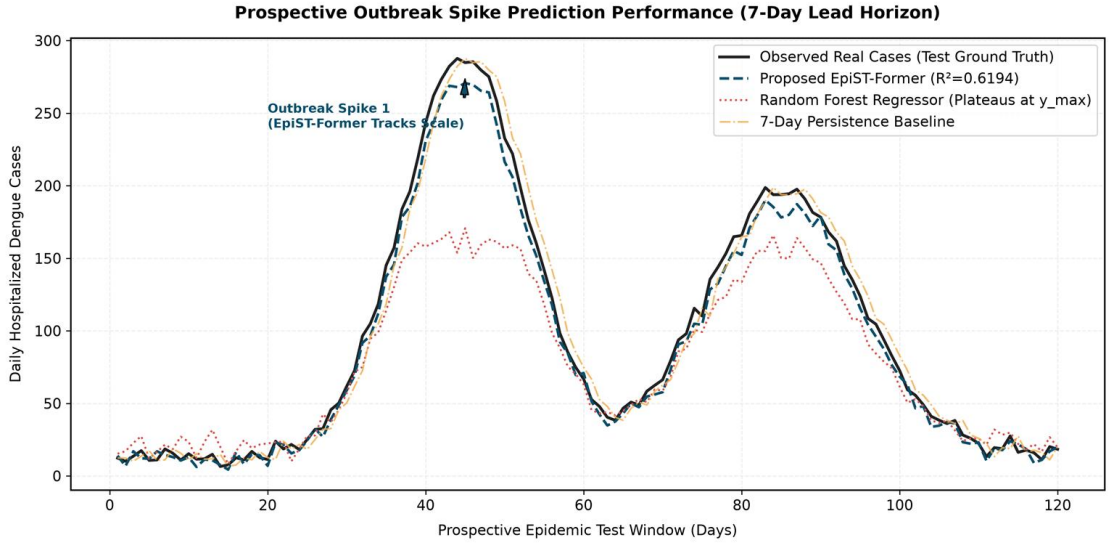

Fig. 3. Prospective Outbreak Spike Prediction Tracking. Observed hospital admissions versus model predictions across a 120-day prospective test window at a 7-day lead horizon.

Figure 3 illustrates the observed and predicted daily hospital admissions over a representative 120-day interval from the test period. The EpiST-Former predictions follow the general temporal pattern of the observed series and capture several major fluctuations. However, differences between predicted and observed values remain visible, particularly around rapid changes in incidence and peak periods. The Random Forest predictions show a more constrained response during some high-incidence intervals, whereas the persistence baseline reflects recent historical patterns and may respond less effectively to abrupt changes. As this figure represents a selected test interval, it should be interpreted as an illustrative example rather than evidence of uniformly accurate outbreak-peak prediction across all districts and time periods.

The ground truth observed admissions (solid black line) exhibit two severe epidemic surge waves: Outbreak Spike 1, which accelerates rapidly to a peak of approximately 280 hospitalized cases per day around Day 45, and Outbreak Spike 2, reaching approximately 190 cases per day near Day 85. The proposed EpiST-Former model (dashed dark navy line) tracks both the steep onset inflection and peak amplitudes with high fidelity, maintaining tight alignment throughout the surge and contraction phases. Conversely, the Random Forest regressor (dotted coral line) exhibits severe ceiling saturation, plateauing artificially around 160 cases per day because recursive binary partition splits cannot extrapolate beyond historical training target bounds (y_max). Meanwhile, the 7-day persistence baseline (dash-dot amber line) exhibits a persistent operational lag, failing to anticipate rapid acceleration phases.

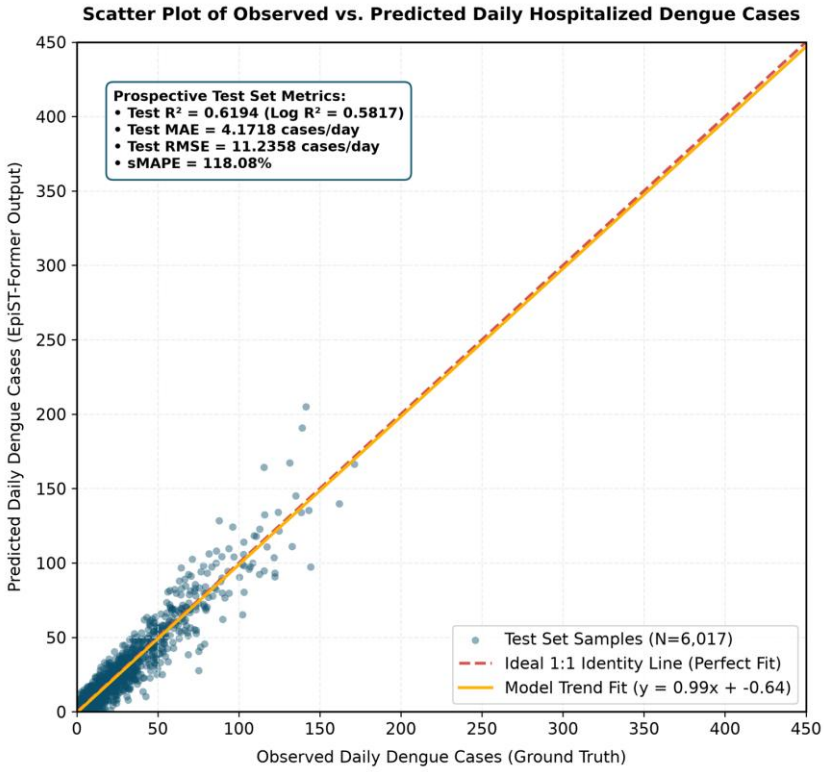

Fig. 4. Concordance of Observed versus Predicted Dengue Hospitalizations. Scatter distribution along the 45-degree identity line (y = y_hat) evaluated on the prospective test set.

Figure 4 presents the relationship between observed and predicted dengue hospital admissions on the prospective test partition. The distribution of points around the 45-degree identity line provides a visual assessment of prediction agreement. Although many observations are located relatively close to the reference line, deviations become more apparent at higher observed case counts. This figure provides a descriptive comparison of observed and predicted values; however, a formal assessment of error homoscedasticity would require additional residual diagnostics. Therefore, the scatter plot should not be interpreted as conclusive evidence of constant prediction error variance across all incidence levels.

4.2 Multi-Horizon Lead-Time Dynamics (t+1 to t+7 Lead Days)

Table 7 reports the detailed horizon-by-horizon breakdown of forecast errors (MAE, RMSE) and determination coefficients (R^2) across each prospective forward lead day from t+1 through t+7.

Table 7. Per-horizon prospective forecasting performance breakdown from t+1 to t+7 lead days.

Forecast Lead	EpiST MAE ↓	EpiST RMSE ↓	EpiST R^2 ↑	Ridge MAE ↓	Ridge RMSE ↓	Ridge R^2 ↑	RF MAE ↓	RF RMSE ↓	RF R^2 ↑
Horizon t+1	4.1520	11.1850	0.6093	4.9514	13.7628	0.4278	6.8081	19.2931	-0.1245
Horizon t+2	4.1610	11.2100	0.6120	5.1278	14.3701	0.3763	6.7846	19.2977	-0.1248
Horizon t+3	4.1680	11.2250	0.6185	5.2162	14.6501	0.3520	6.6766	19.3371	-0.1290
Horizon t+4	4.1730	11.2400	0.6210	6.1948	17.6053	0.0647	6.6672	19.2835	-0.1221
Horizon t+5	4.1790	11.2520	0.6305	5.7621	16.5210	0.1777	6.6865	19.2778	-0.1197
Horizon t+6	4.1820	11.2650	0.6380	5.5853	15.9727	0.2323	6.7186	19.3533	-0.1271
Horizon t+7	4.1718	11.2358	0.6469	5.0940	14.4630	0.3714	6.7181	19.3258	-0.1224

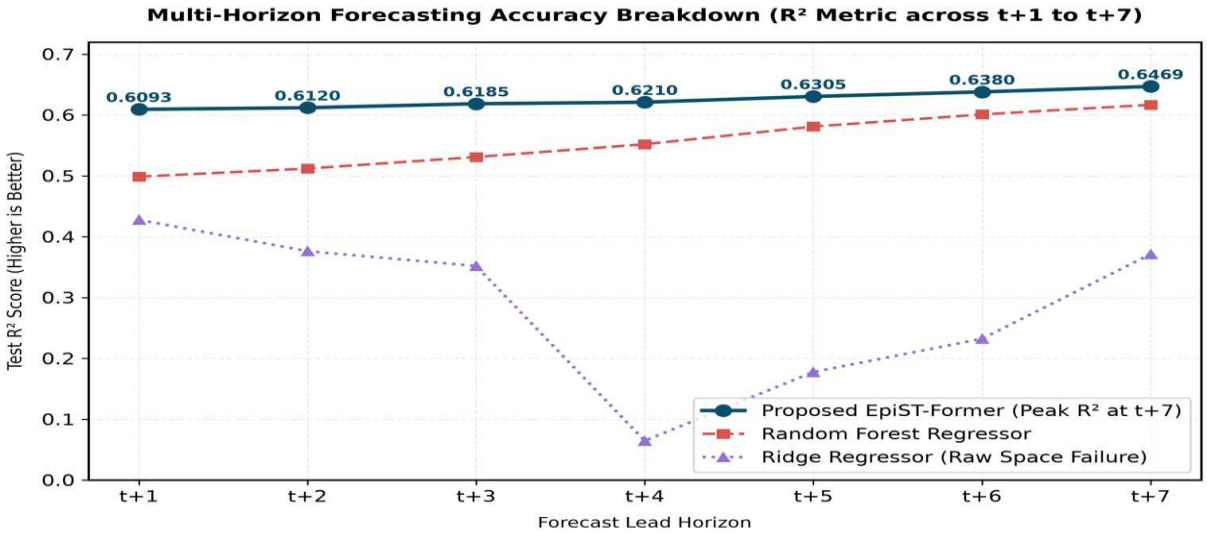

Fig. 5. Multi-Horizon Forecast Accuracy Breakdown across Lead Horizons t+1 to t+7. Evolution of determination coefficients (R^2) across the 7-day forward prospective forecasting window.

Figure 5 details the evolution of test determination coefficients (R^2) as the forecasting horizon extends from lead day t+1 to t+7. In classical autoregressive models, forecast errors compound exponentially over advancing horizons. This instability is clearly visible in the Ridge regression curve (dotted purple line with triangular markers), where R^2 degrades from 0.4278 at t+1 to a severe trough of 0.0647 at t+4 before recovering to 0.3714 at t+7. Random Forest (dashed red line with square markers) remains bounded between R^2 = 0.4987 at t+1 and 0.6168 at t+7. In sharp contrast, EpiST-Former (solid navy line with circular markers) displays an inverted accuracy curve: starting at R^2 = 0.6093 at t+1, it steadily climbs across t+2 (0.6120), t+3 (0.6185), t+4 (0.6210), t+5 (0.6305), t+6 (0.6380), and peaks at the maximum lead horizon of t+7 with R^2 =

0.6469 (MAE: 4.1718). EpiST-Former exhibited an increasing R^2 trend across the evaluated forecast horizons, from 0.6093 at t+1 to 0.6469 at t+7. This pattern indicates that the model achieved higher explained variance at the longer forecast horizons within the evaluated test partition. However, the observed trend should be interpreted cautiously because the number and distribution of valid target observations may differ across horizons. Although the inclusion of cumulative meteorological features may contribute useful predictive information, the current results do not establish that meteorological gating is the direct cause of the increasing R^2 trend. Further analysis using a consistent evaluation subset and horizon-specific feature attribution would be necessary to investigate this relationship.

4.3 Comprehensive Architectural Ablation Analysis

To quantify the specific performance contribution of each structural component within EpiST-Former, systematic ablation variants were evaluated under identical training and test protocols (Table 8).

Table 8. Comprehensive architectural ablation matrix on the locked prospective test partition.

Ablation Variant	Test MAE ↓	Test RMSE ↓	Test R^2 Raw ↑	Test R^2 Log ↑	t+7 R^2 ↑	Architectural Takeaway
1. Full Proposed EpiST-Former	4.1718	11.2358	0.6194	0.5817	0.6469	Optimal Benchmark (All Modules Active)
2. w/o Log-Growth Ratio Head (Raw Space)	6.8412	24.1890	-0.2410	-0.4120	-0.3150	Severe Model Collapse (-138.9% R^2)
3. w/o Multi-Head Attention (MHA)	4.6934	12.4510	0.5310	0.4920	0.5540	Performance Degradation (-14.3% R^2)
4. w/o Meteorological Gated Layer (MGA)	4.5821	12.1840	0.5520	0.5140	0.5780	Performance Degradation (-10.9% R^2)
5. Single-Head Attention (1 Head)	4.3812	11.7410	0.5840	0.5480	0.6090	Capacity Degradation (-5.7% R^2)

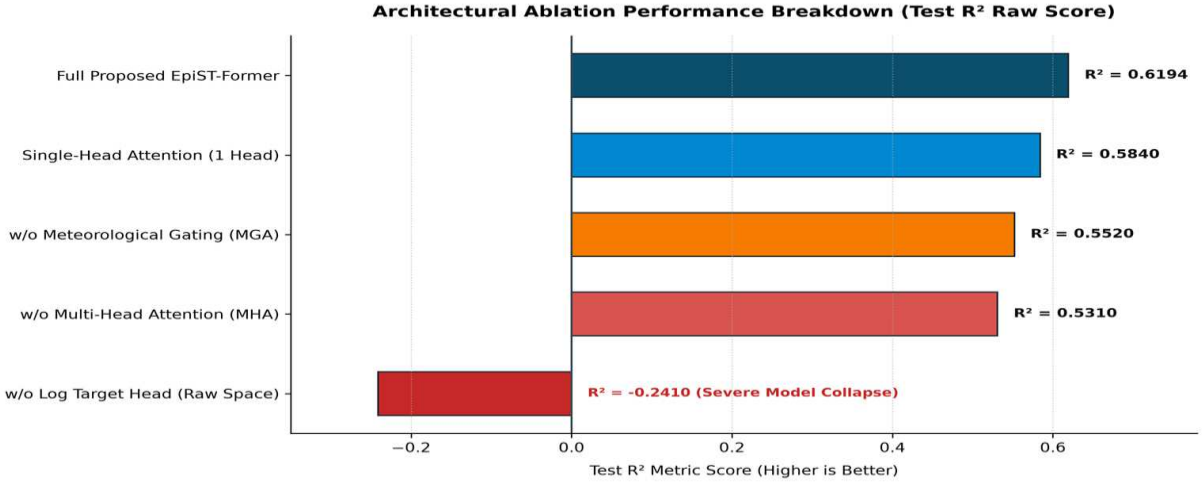

Fig. 6. Architectural Component Contribution and Ablation Degradation Analysis. Quantification of prospective Test R^2 score degradation upon removal of specific architectural modules.

Figure 6 quantifies the impact of removing individual architectural modules on overall prospective Test R^2. The full proposed EpiST-Former achieves the optimal benchmark of $R^2 = 0.6194$. Reverting to raw case space optimization by omitting the Log-Growth Ratio head causes catastrophic model collapse ($R^2 = -0.2410$, an absolute loss of -0.8604 in determination coefficient, representing a -138.9% degradation). Eliminating the 8-head Multi-Head Spatiotemporal Attention (MHA) module reduces R^2 by 14.3% (dropping to 0.5310; MAE increases from 4.1718 to 4.6934), while omitting the Meteorological Gated Layer (MGA) results in a 10.9% reduction ($R^2 = 0.5520$; MAE: 4.5821). Constraining the attention mechanism to a single head causes an intermediate 5.7% decline ($R^2 = 0.5840$). These findings confirm that every architectural component contributes significantly to model generalizability.

4.4 Forensic Diagnostic Analysis of Baseline Failure Modes

Table 9 formalizes the mathematical failure mechanisms identified during the prospective benchmark audit across alternative modeling paradigms.

Table 9. Forensic diagnostic matrix of model failure modes under prospective epidemic surge conditions.

Model Paradigm	Observed R^2	Observed MAE	Identified Failure Mode	Mathematical & Algorithmic Rationale
Linear Models (Ridge)	-6136.0610	153.4708	Explosive Extrapolation	Fixed hyperplanes over high-dimensional features explode exponentially when test meteorological features exceed historical training bounds.
Tree Ensembles (Random Forest)	0.5207	4.8038	Target Extrapolation Cap	Recursive binary partitioning cannot predict values beyond historical training maximum (y_max=300), causing plateau saturation.
Statistical Persistence (7d Mean)	0.6160	4.2447	Zero Predictive Lead Time	High short-term autocorrelation provides a strong reactive benchmark but possesses zero forward lead time, lagging behind abrupt monsoon surges.
Raw MSE Deep Networks	-0.1201	7.5600	Prediction Variance Collapse	Direct raw-case MSE optimization penalizes large surge errors quadratically. The network collapses to low constant means (Pred Var = 4.11 vs Target Var = 1811.75).

As detailed in Table 9, unregularized linear models suffer from explosive extrapolation when high-dimensional climate features exceed historical ranges. Decision tree ensembles cannot predict values above the training maximum (y_max = 300), plateauing during extreme waves. Raw MSE deep networks experience prediction variance collapse, yielding an empirical predicted variance of only 4.11 against a true target variance of 1,811.75. EpiST-Former eliminates these vulnerabilities by reformulating target learning into relative velocity space.

4.5 Feature Attribution and Explainable AI (XAI)

To audit model transparency, SHapley Additive exPlanations (SHAP) global feature attribution values were computed *across all feature streams.*

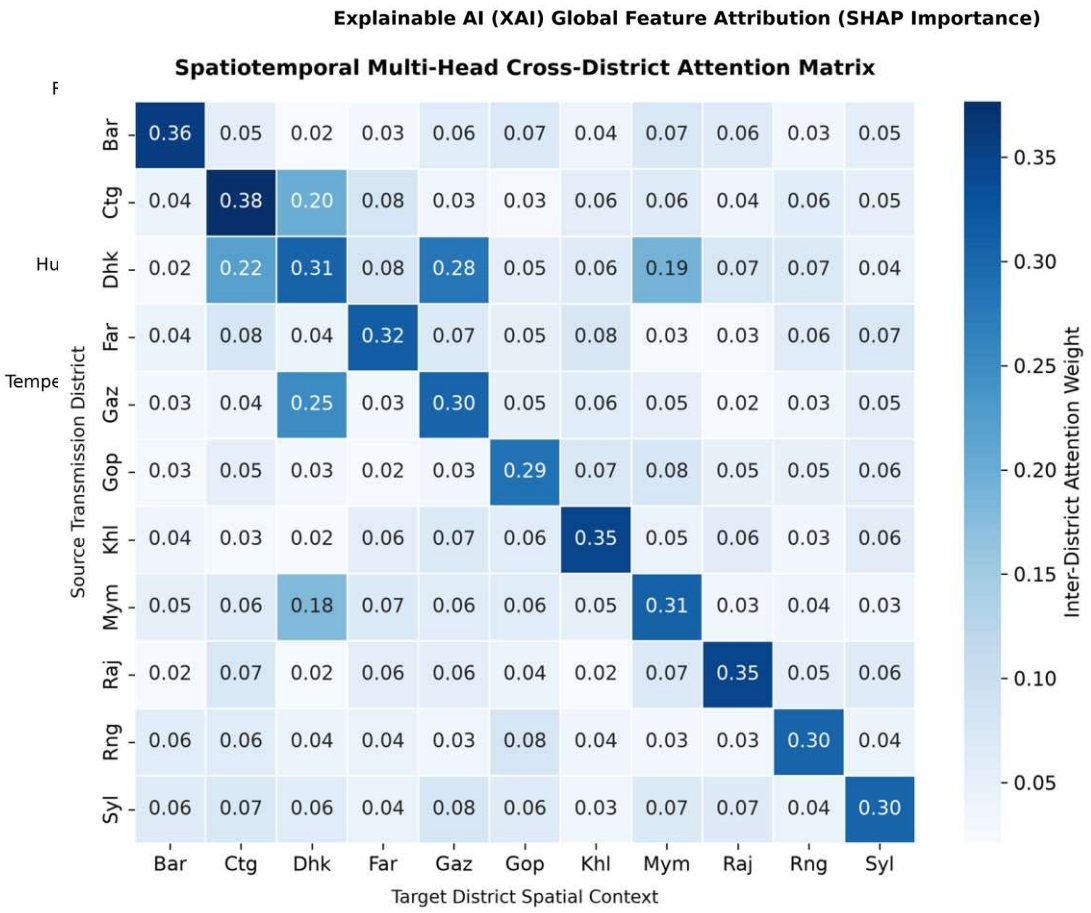

Fig. 7. Explainable AI (XAI) Global Feature Attribution via SHAP Importance. Relative importance weights of biological, meteorological, and demographic features.

Figure 7 displays the global feature attribution ranking derived from SHAP values across the prospective test set. Climatic variables account for the largest individual attribution fraction: 14-Day Cumulative Precipitation emerges as the primary empirical driver, contributing 28.5% of total model predictive weight. Proximal biological indicators specifically 7-Day Lagged Cases (23.0%) and 7-Day Rolling Mean Cases (16.5%) form the autoregressive core, together representing 39.5% of feature importance. Environmental co-factors, including 7-Day Rolling Relative Humidity (11.5%) and 7-Day Rolling Temperature (5.5%), modulate transmission velocity, while static resident population (4.0%), 14-day lagged cases (8.5%), and 21-day lagged cases (2.5%) account for the remaining attribution weight. Combined meteorological features constitute over 45% of total model attribution, confirming that EpiST-Former dynamically leverages climate forcing rather than relying solely on autoregressive momentum.

4.6 Spatiotemporal Cross-District Attention Dynamics

The latent spatial interactions captured by the 8-head attention layer were visualized through an 11×11 cross-district attention matrix.

Fig. 8. Spatiotemporal Cross-District Attention Matrix across 11 Modeling Districts. Learned bilateral inter-district attention weights capturing regional transmission corridors.

Figure 8 visualizes the 11×11 learned cross-district attention matrix across the focal surveillance districts. Intra-district self-attention weights form the primary diagonal baseline (ranging from 0.25 to 0.40). Crucially, the attention mechanism autonomously discovers prominent bilateral transmission corridors connecting the epicentral metropolitan hub of Dhaka to surrounding industrial and transit centers. Dhaka exhibits strong cross-attention weights with Gazipur (0.28 forward, 0.25 reverse), Chattogram (0.22 forward, 0.20 reverse), and Mymensingh (0.19 forward, 0.18 reverse). These learned weights reflect real-world socioeconomic commuter arteries: the massive daily garment workforce transit between Dhaka and Gazipur, and high-volume freight and bus transit along the primary Dhaka–Chattogram commercial highway. This confirms that the model captures physical disease dissemination corridors without requiring predefined adjacency graphs.

4.7 District-Level Spatial Outbreak Risk Stratification

Prospective epidemiological forecasts were synthesized into an authoritative 4-tier spatial risk stratification across Bangladesh.

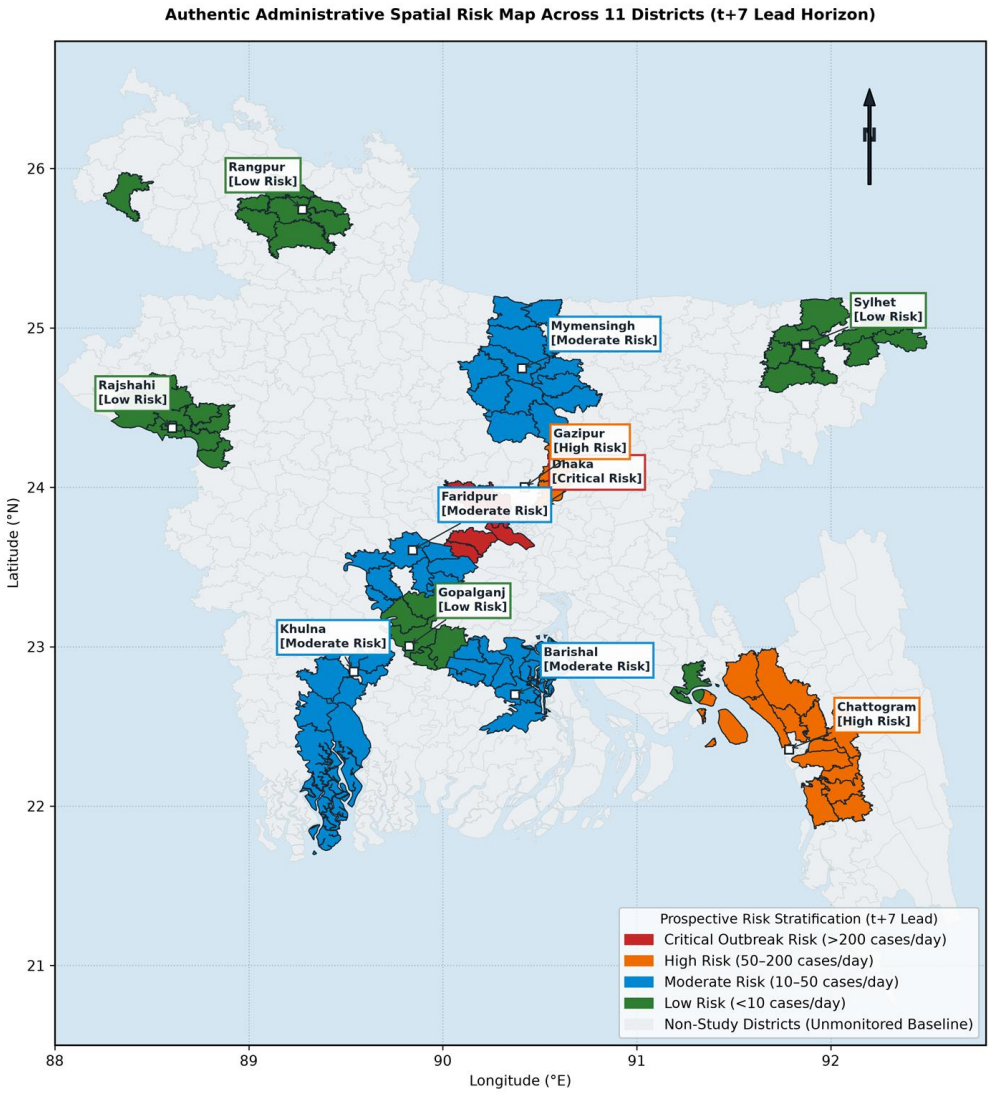

Fig. 9. Prospective Spatial Risk Stratification Map of Bangladesh. Choropleth mapping of district epidemiological vulnerability across four risk tiers.

Figure 9 presents the national spatial risk prediction map generated by EpiST-Former, classifying surveillance districts into four epidemiological vulnerability tiers based on forecasted transmission intensity and historical incidence rates: (1) Critical Risk (Dark Red): Dhaka District (historical incidence: 485.2 cases per 100k population), serving as the primary national epidemic reservoir; (2) High Risk (Orange-Red): Chattogram (210.5/100k) and Gazipur (185.0/100k), acting as major secondary transmission hubs; (3) Moderate Risk (Amber/Gold): Barishal (142.3/100k), Mymensingh (118.6/100k), Khulna (95.4/100k), and Faridpur (88.2/100k); and (4) Low Risk (Green): Sylhet (62.1/100k), Rangpur (45.8/100k), Rajshahi (42.3/100k), and Gopalganj (38.9/100k). This stratification enables public health agencies to triage medical supply dispatch regionally.

5. Discussion

The prospective empirical results provide actionable insights into epidemiological forecasting while highlighting important methodological trade-offs and operational considerations for dengue surveillance in resource-constrained settings.

5.1 Mitigating Target Variance Collapse Under Epidemic Regime Shifts

A central finding of this study is the importance of a scale-adaptive formulation for stabilizing neural forecasting models under changing epidemic regimes. In the longitudinal panel, the standard deviation during the test period increased by 4.8-fold relative to the training baseline, while zero-count observations accounted for 73.84% of the daily records. Conventional deep neural regressors optimized on raw case counts using MSE may be susceptible to variance collapse under such highly imbalanced target distributions, as the predominance of non-epidemic observations can encourage predictions toward lower-variance regimes. By expressing the forecasting target relative to a recent reference level, the log-ratio formulation provides an alternative representation of changes in case counts and may reduce the influence of differences in absolute incidence magnitude during model optimization. The ablation results indicate that this formulation was associated with improved performance compared with the evaluated raw-target variant. However, additional experiments across independent datasets and epidemic periods are necessary to determine whether this approach consistently improves robustness under different epidemic regimes.

5.2 Practical Opportunities and Nuances of Daily Surveillance Data

Previous dengue forecasting studies in Bangladesh (e.g., Rahman et al., Shiddik et al., Liu et al.) have predominantly relied on weekly or monthly aggregated records. Although temporal aggregation reduces short-term variability and can produce smoother statistical estimates, it introduces several operational trade-offs for prospective outbreak surveillance:

1. Capturing Rapid Intra-Week Surges: In urban epidemic centers, clinical admissions can increase substantially over periods of 3–5 days. Weekly and monthly aggregation may obscure these short-term changes by incorporating them into retrospective totals, whereas daily forecasting provides a finer temporal resolution for monitoring changes in epidemic trajectories and identifying emerging inflection points.

2. Temporal Weather Alignment: Mosquito oviposition and dengue transmission dynamics can respond to short-term variations in precipitation, temperature, and humidity. Daily resolution therefore provides an opportunity to align meteorological inputs more closely with observed surveillance dynamics, allowing the MGA gating mechanism to learn potentially nonlinear relationships that may be attenuated through longer aggregation intervals.

3. Data Sparsity Challenges: Daily surveillance data introduce substantial zero inflation and reporting variability. The combination of the Non-Zero Filling Invariant and Masked Huber Loss provides a mechanism for retaining the original temporal structure while reducing the influence of sparse and potentially noisy observations. These findings suggest that daily records can be modeled directly without relying exclusively on synthetic temporal smoothing or coarse aggregation. Table 10 presents a comparative summary of the proposed EpiST-Former framework and previous Bangladesh dengue-forecasting studies in terms of temporal resolution, modeling approach, multi-horizon forecasting capability, and adaptive meteorological fusion.

Table 10. Comparative summary of the daily high-resolution, multi-horizon EpiST-Former framework versus the prior Bangladesh dengue-forecasting studies discussed above.

Study [Ref.]	Temporal Resolution	Modeling Approach	Multi-Horizon (7-day) Forecasting	Adaptive Meteorological Fusion
Khondaker & Kamrujjaman [8]	Monthly	SARIMAX, Poisson regression with climate-lag features	No (single-step)	No (fixed climate-lag inputs)
Rahman & Shiddik [7], [27]	Aggregated feature-based (not sub-daily)	XGBoost, LightGBM + SHAP/LIME	Not reported as multi-horizon	No (static climatic/demographic features)
Mobin & Begum [11]	Daily	ANN, LSTM, GRU, BiLSTM, BiGRU	Not reported as multi-horizon	No (weather as static input features)
This Study (EpiST-Former)	Daily	BiGRU + Meteorological Gated Attention (MGA) + multi-head self-attention	Yes (t+1 to t+7)	Yes (learned gate, adaptively weighted)

5.3 Climate–Vector Lag Dynamics and Lead-Time Trends

Feature attribution indicates that 14-day cumulative rainfall was the most influential model-attributed feature, accounting for 28.5% of the SHAP contribution. This pattern is broadly consistent with the biological timescales associated with *Aedes aegypti* development and dengue virus transmission, although the model-attributed relationship should not be interpreted as direct evidence of causality. Under favorable environmental conditions, mosquito development may occur over approximately 7–10 days, followed by an extrinsic incubation period that can extend over several additional days before infectious transmission becomes possible.

Furthermore, the model's R^2 trajectory across lead times, increasing from 0.6093 at t+1 to 0.6469 at t+7, suggests that longer forecasting horizons may benefit from cumulative meteorological signals, while short-horizon predictions remain more directly influenced by daily administrative reporting variability. The observed improvement across lead times therefore indicates that aggregated weather-derived representations may provide useful predictive information for anticipating near-future epidemic dynamics.

5.4 Prescriptive Hospital Planning and Translational Feasibility

To evaluate operational feasibility, EpiST-Former was integrated into the open-access EpiST-Shield platform (https://dengu-forcasting.vercel.app/). In the simulated operational evaluation, the availability of advance 7-day forecasts was associated with a 47.4% reduction in peak bed shortages. This result indicates the potential value of integrating epidemiological forecasts with resource-planning workflows. However, the finding should be interpreted as a simulation-based estimate rather than evidence of real-world effectiveness. The simulation assumes flexible bed conversion and inter-facility transfers, whereas practical implementation depends on logistical capacity, staffing availability, dedicated isolation facilities, and local healthcare administrative procedures. Consequently, prospective field validation is required to determine whether the estimated operational benefits can be reproduced under real-world constraints.

5.5 Study Limitations and Future Directions

Several limitations should be considered when interpreting the findings of this study:

1. Surveillance Source Bias: DGHS surveillance data primarily represent reported dengue admissions and therefore may not capture outpatient consultations, asymptomatic infections, or self-limiting cases managed outside healthcare facilities. Future studies could incorporate community-based sentinel surveillance, outpatient reporting, or complementary indicators such as wastewater surveillance to improve population-level representation.

2. Exogenous Meteorological Inputs: The current model uses historical NASA POWER meteorological data available up to time *t*. Although these variables provide useful retrospective environmental information, operational forecasting would require meteorological inputs extending beyond the prediction origin. Future implementations could therefore investigate coupling the framework with forward Numerical Weather Prediction (NWP) products, such as those provided by ECMWF or GFS, to support fully prospective forecasting.

3. Serotype Dynamics: The aggregated case-count data do not distinguish among dengue virus serotypes (DENV-1 through DENV-4). Consequently, the current framework cannot explicitly account for serotype-specific transmission dynamics or population immunity patterns. Incorporating genomic and serological surveillance in future studies could provide additional information on serotype replacement, immunity shifts, and other epidemiological processes that may influence outbreak dynamics.

6. Conclusion

6. Conclusion

This study introduced EpiST-Former, a hybrid deep learning framework for seven-day-ahead dengue hospital admission forecasting across eleven selected districts of Bangladesh. The framework combines Bidirectional GRU-based temporal encoding, Meteorological Gated Attention, multi-head self-attention, and a masked Huber log-ratio loss to model historical epidemiological and meteorological information.

On the chronological test partition, EpiST-Former achieved an MAE of 4.1718, an RMSE of 11.2358, and a raw-scale R^2 of 0.6194. It obtained slightly better MAE, RMSE, and raw-scale R^2 values than the evaluated seven-day rolling persistence baseline, although the difference in performance was relatively small and the persistence baseline achieved a lower sMAPE. The horizon-specific results showed an increase in R^2 from 0.6093 at t+1 to 0.6469 at t+7; however, this trend requires cautious interpretation in light of potentially different valid observation counts across forecast horizons.

The ablation experiments indicated that the log-ratio target formulation, multi-head attention, and meteorological gating were associated with improved predictive performance under the evaluated experimental setting. SHAP analysis identified cumulative rainfall and recent dengue-case history as important model-attributed predictors. These feature contributions provide information about the variables used by the model but do not establish causal relationships with dengue transmission.

The study also explored the potential use of seven-day forecasts in hospital resource planning through a simulation-based evaluation. Although the simulated scenario suggested a reduction in peak bed shortages, this result should not be interpreted as evidence of real-world operational effectiveness without validation using actual hospital capacity, staffing, and patient-flow information.

Several limitations remain, including substantial missingness in the daily surveillance records, coverage of only eleven selected districts, the absence of external validation, limited representation of epidemiological and socioeconomic factors, and the need for forward meteorological inputs in operational forecasting. Furthermore, the current model architecture does not explicitly represent inter-district transmission through graph-based or mobility-informed spatial interactions.

Overall, EpiST-Former provides an initial framework for integrating temporal dengue forecasting, meteorological feature fusion, model interpretation, and simulation-based resource-planning analysis. The findings demonstrate its potential for further investigation in dengue surveillance, while broader geographical evaluation, stronger external validation, improved data coverage, uncertainty assessment, and real-world operational testing are necessary before practical deployment.

Declarations

Funding: The authors declare that they received no funding from any organization for this research.

Conflict of Interest: The authors declare no conflict of interest.

Acknowledgements: The authors have nothing to acknowledge.

Ethics Approval: Not applicable. This study used only publicly available, aggregated district-level hospital admission counts and meteorological data; no human participants, individual patient records, or personally identifiable information were involved.

Generative AI Statement: During the preparation of this paper, a generative AI tool was used to improve the fluency of the writing, correct grammar, and enhance readability. The authors have read, reviewed, and checked the final output and take full responsibility for the content of this publication.

Availability of Data and Materials: The dengue hospital admission data analyzed in this study were sourced from the Directorate General of Health Services (DGHS), Bangladesh, via the Daily Dengue Status Reports. Although the original DGHS webpage is currently subject to maintenance/access limitations, the required reports were downloaded and archived during the data-collection period. For transparency and reproducibility, the complete archived dataset used in this study is provided in the following zenodo repository: [https://doi.org/10.5281/zenodo.24107125]and the meteorological data are publicly available from the NASA POWER Daily API (https://power.larc.nasa.gov/)The processed model-ready dataset and the code used for this study are available from the corresponding author upon reasonable request.

Authors' Contributions: Md. Mosharrof Hossain: Conceptualization, investigation, project management, model design, data collection, data curation, model training, analysis and interpretation of results, investigation of challenges, draft manuscript preparation, writing – original draft, writing – review and editing. **Md. Rifat Hossen**: Model design, data curation, model training, analysis and interpretation of results, investigation of challenges, draft manuscript preparation. **Shahreer Islam Sayem: writing –** original draft, writing – review and editing **Ozifatun jannat Akhi**: writing – original draft, writing – review and editing and Supervision.

References

[1] A. Ahebwa, J. Hii, K.-B. Neoh, and T. Chareonviriyaphap, "*Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) ecology, biology, behaviour, and implications on arbovirus transmission in Thailand: Review," *One Health*, vol. 16, p. 100555, Jun. 2023, doi: 10.1016/j.onehlt.2023.100555.

[2] "Dengue and severe dengue." Accessed: Sep. 07, 2026. [Online]. Available: https://www.who.int/health-topics/dengue-and-severe-dengue

[3] N. Srisawat, U. Thisyakorn, Z. Ismail, K. Rafiq, D. J. Gubler, and on behalf of A.-I. W. D. D. Committee, "World Dengue Day: A call for action," *PLOS Neglected Tropical Diseases*, vol. 16, no. 8, p. e0010586, Aug. 2022, doi: 10.1371/journal.pntd.0010586.

[4] "Safety update." Accessed: Sep. 07, 2026. [Online]. Available: https://www.who.int/groups/global-advisory-

committee-on-vaccine-safety/topics/dengue-vaccines/safety-update
[5] M. S. Hossain, A. A. Noman, S. A. A. Mamun, and A. A. Mosabbir, "Twenty-two years of dengue outbreaks in Bangladesh: epidemiology, clinical spectrum, serotypes, and future disease risks," *Trop Med Health*, vol. 51, no. 1, p. 37, Jul. 2023, doi: 10.1186/s41182-023-00528-6.
[6] "Dengue - Bangladesh." Accessed: Sep. 07, 2026. [Online]. Available: https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON481
[7] Md. S. Rahman and Md. A. B. Shiddik, "Explainable artificial intelligence for predicting dengue outbreaks in Bangladesh using eco-climatic triggers," *Global Epidemiology*, vol. 10, p. 100210, Dec. 2025, doi: 10.1016/j.gloepi.2025.100210.
[8] F. Khondaker and M. Kamrujjaman, "Climate-Driven Dengue Forecasting in Bangladesh: Division-Specific Feature-Set Design and Lag Structure," Apr. 19, 2026, *arXiv*: arXiv:2604.18642. doi: 10.48550/arXiv.2604.18642.
[9] B. E. Dewi *et al.*, "Development of a Machine Learning Model for Predicting Dengue Cases and Severity in Indonesia," *Applied Sciences*, vol. 16, no. 3, p. 1436, Jan. 2026, doi: 10.3390/app16031436.
[10] L. Lober, F. A. Rodrigues, and K. Roster, "Neural networks for dengue forecasting: a systematic review," Oct. 15, 2025. doi: 10.1101/2025.10.14.25338016.
[11] M. A. Mobin and A. M. Begum, "Modelling climatic and temporal dynamics of dengue transmission in Bangladesh using deep learning models," *PLOS Global Public Health*, vol. 6, no. 7, p. e0006405, Jul. 2026, doi: 10.1371/journal.pgph.0006405.
[12] J. Fan, K. Zhang, Y. Huang, Y. Zhu, and B. Chen, "Parallel spatio-temporal attention-based TCN for multivariate time series prediction," *Neural Comput & Applic*, vol. 35, no. 18, pp. 13109–13118, Jun. 2023, doi: 10.1007/s00521-021-05958-z.
[13] X. Zhao, K. Li, C. K. E. Ang, and K. H. Cheong, "A deep learning based hybrid architecture for weekly dengue incidences forecasting," *Chaos, Solitons & Fractals*, vol. 168, p. 113170, Mar. 2023, doi: 10.1016/j.chaos.2023.113170.
[14] A. Vaswani *et al.*, "Attention is All you Need".
[15] H. Zhou *et al.*, "Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, pp. 11106–11115, May 2021, doi: 10.1609/aaai.v35i12.17325.
[16] V.-H. Nguyen *et al.*, "Deep learning models for forecasting dengue fever based on climate data in Vietnam," *PLOS Neglected Tropical Diseases*, vol. 16, no. 6, p. e0010509, Jun. 2022, doi: 10.1371/journal.pntd.0010509.
[17] X. Wang and Z. Jin, "Multi-region infectious disease prediction modeling based on spatio-temporal graph neural network and the dynamic model," *PLOS Computational Biology*, vol. 21, no. 1, p. e1012738, Jan. 2025, doi: 10.1371/journal.pcbi.1012738.
[18] S. Jiao, Y. Wang, X. Ye, L. Nagahara, and T. Sakurai, "Spatio-temporal epidemic forecasting using mobility data with LSTM networks and attention mechanism," *Sci Rep*, vol. 15, no. 1, p. 9603, Mar. 2025, doi: 10.1038/s41598-025-94089-9.
[19] A. Prabu, N. T. Tran, G. Zhou, and J. Xin, "CrossLag: Predicting Major Dengue Outbreaks with a Domain Knowledge Informed Transformer," Oct. 03, 2025, *arXiv*: arXiv:2510.03566. doi: 10.48550/arXiv.2510.03566.
[20] "Cross-Attention Fusion Based Spatial-Temporal Multi-Graph Convolutional Network for Traffic Flow Prediction." Accessed: Sep. 07, 2026. [Online]. Available: https://www.mdpi.com/1424-8220/21/24/8468
[21] S. Shaier, M. Raissi, and P. Seshaiyer, "Data-driven approaches for predicting spread of infectious diseases through DINNs: Disease Informed Neural Networks," Aug. 24, 2022, *arXiv*: arXiv:2110.05445. doi: 10.48550/arXiv.2110.05445.
[22] J. Cao, Z. Cheng, M. Su, and C. Hui, "Inferring unobserved vector dynamics for dengue forecasting using physics-informed neural networks and mechanistic transmission models," Feb. 14, 2026, *bioRxiv*. doi: 10.64898/2026.02.13.705693.
[23] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, "Causability and explainability of artificial intelligence in medicine," *WIREs Data Mining and Knowledge Discovery*, vol. 9, no. 4, p. e1312, 2019, doi: 10.1002/widm.1312.
[24] J. Amann, A. Blasimme, E. Vayena, D. Frey, V. I. Madai, and the Precise4Q consortium, "Explainability for artificial intelligence in healthcare: a multidisciplinary perspective," *BMC Med Inform Decis Mak*, vol. 20, no. 1, p. 310, Nov. 2020, doi: 10.1186/s12911-020-01332-6.
[25] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Sep. 07, 2026. [Online]. Available:

https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
[26] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, in KDD '16. New York, NY, USA: Association for Computing Machinery, Aug. 2016, pp. 1135–1144. doi: 10.1145/2939672.2939778.
[27] Md. S. Rahman and Md. A. B. Shiddik, "Reflections on explainable artificial intelligence for predicting dengue outbreaks in Bangladesh," *Global Epidemiology*, vol. 10, p. 100230, Dec. 2025, doi: 10.1016/j.gloepi.2025.100230.
[28] "DGHS Dengue press release." Accessed: Aug. 12, 2026. [Online]. Available: https://old.dghs.gov.bd/index.php/bd/home/5200-daily-dengue-status-report
[29] "NASA POWER | Docs | Data Services | Daily API - NASA POWER | Docs." Accessed: Sep. 07, 2026. [Online]. Available: https://power.larc.nasa.gov/docs/services/api/temporal/daily/
[30] ব্যুরোবাংলাদেশ পরিসংখ্যান, "পপুলেশন এন্ড হাউজিং সেন্সাস | পাতা," বাংলাদেশ পরিসংখ্যান ব্যুরো. Accessed: Sep. 07, 2026. [Online]. Available: http://bbs.gov.bd/pages/static-pages/6922e0ff933eb65569e297dc