SpaTempNet: Deep Spatiotemporal Forecasting of River Morphological Evolution of the Padma River in Bangladesh

S. S. Mahmud Turza[1], Md. Kaoser Ahamed Anik[2], Md. Shadmim Hasan Sifat[3],
Mahruba Sharmin Chowdhury[4], Khandokar Md. Rahat Hossain[5], Abdullah Al Noman[6]
Dept. of Computer Science and Engineering, Shahjalal University of Science and Technology, Sylhet, Bangladesh[1,2,3,4]
Dept. of Computer Science and Engineering, United International University, Dhaka, Bangladesh[5]
Computer Science and Engineering Discipline, Khulna University, Khulna 9208, Bangladesh[6]

Abstract*—Rivers are dynamic geomorphological systems experiencing continuous transformation through erosion and sediment transport. In Bangladesh, these processes displace 50,000-200,000 people annually, causing severe losses along the Padma River according to Natural Resources Defense Council (NRDC). Traditional monitoring via field surveys and manual remote sensing proves costly, spatially limited, and unsuitable for multi-year prediction. This study presents a comprehensive framework, termed SpaTempNet, for forecasting river morphological evolution from freely available satellite data using spatiotemporal deep learning. A 38-year time-series (1987–2025) of binary water masks was constructed from Landsat (5 TM, 7 ETM+, 8 OLI) and Sentinel (Sentinel-2 MSI, Sentinel-1 SAR) imagery via Google Earth Engine. Water extraction used Modified Normalised Difference Water Index (MNDWI). The proposed Bidirectional ConvLSTM gap-filling model achieved IoU = 0.7366, outperforming classical methods. Five spatiotemporal architectures (ConvLSTM, U-Net+LSTM, Attention U-Net+ConvLSTM, Swin Transformer, ViT-based model) were evaluated across yearly, quarterly, and bi-monthly resolutions. Hybrid CNN-LSTM models consistently outperformed pure transformers. Attention U-Net+ConvLSTM achieved best yearly performance (IoU = 0.7005); U-Net+LSTM led bi-monthly prediction (IoU = 0.7791). Statistical analysis quantified mean annual migration of 255.4 m yr^{-1} with 1998 extreme of 1,485.5 m yr^{-1}. An expansion of about 290 km^2 was projected, and a spatially explicit risk map was generated through long-term forecasting (2026-2040). Results demonstrate that freely available satellite imagery with deep learning can provide practical, scalable framework for riverbank hazard monitoring and disaster management.

Keywords*—*River morphology; riverbank erosion forecasting; spatiotemporal neural networks; ConvLSTM; U-Net; Swin Transformer; Vision Transformer; satellite remote sensing; MNDWI; disaster management

I. Introduction

Rivers are one of the most morphologically dynamic features that characterize the terrestrial surface of the earth. Because of continuous fluvial erosion, sediment transport, and depositional accretion, they undergo constant geomorphological reshaping. These processes occur in broad spatial and temporal scales, from daily fluctuations during floods to decadal shifts in channel position, planform geometry, and cross-sectional morphology. Comprehensive understanding and reliable predictive capability regarding these morphodynamic transformations are fundamental to effectively integrate water resource management, safeguard human settlements and infrastructure, maintain agricultural productivity in riparian zones, and preserve critical riparian and aquatic ecosystem integrity and biodiversity [1].

Bangladesh stands on a risky geomorphological and hydrological position within this global context. It occupies the terminal deltaic sink of the integrated Ganges-Brahmaputra-Meghna river system, which is regarded as one of Earth's most extensive active delta complexes. Its river network shows exceptional suspended sediment loads, which exceeds 1 billion tonnes annually, and intense seasonal flooding- driven by South Asian monsoon meteorological forcing [2], [3]. The region experiences some of the highest lateral channel migration rates recorded globally, often exceeds several hundred metres per monsoon season and occasionally reaches kilometre-scale adjustments. These extreme morphodynamic processes cause persistent and severe riverbank erosion that systematically displaces between 50,000 and 200,000 individuals annually [4]. The consequences are devastating cumulative economic losses through destruction of agricultural land, transportation infrastructure (roads, bridges, ferries), flood protection embankments, educational facilities, and healthcare infrastructure. Beyond physical infrastructure, erosion destroys the fundamental social fabric binding riverine communities. The socioeconomic consequences extend far beyond immediate physical displacement. Affected populations face long-term food insecurity, disrupted educational access for children, dissolution of social networks and community structures, and profoundly constrained livelihood trajectories and economic opportunities [3], [5], [6]. The most economically vulnerable households bear disproportionate burdens, lacking resources for adaptation or relocation.

Despite more than five decades of continuously accumulated satellite remote sensing imagery archives now being comprehensively available through multiple international space agencies and data repositories, these invaluable spatiotemporal datasets remain substantially under-exploited for predictive morphological modelling applications beyond retrospective change detection. The dominant research paradigm characterizing existing peer-reviewed literature emphasizes retrospective change detection methodologies: quantitative comparison of discrete temporal snapshots to characterize and quantify morphological transformations that have already occurred, rather than developing prospective forecasting capabilities to predict transformations that will occur. Classical remote sensing analytical approaches inherently overlook the continuous temporal evolution trajectories and complex spatial dependencies intrinsically characterizing river morphodynamic processes. Recently some transformative breakthroughs are seen in deep spatiotemporal learning architectures, particularly in computational frameworks. Domains include convolutional spatial feature extraction and recurrent temporal memory mechanisms. These have demonstrated unprecedented success in learning complex sequential patterns from high-dimensional image sequence data [7], [8]. However, systematic application of these powerful architectures specifically to multi-year river morphology spatial forecasting remains comparatively sparse within published literature. Also, comprehensive comparative evaluation rigorously assessing multiple competing architectures for this specific forecasting task under controlled experimental conditions is rare.

Based on the existing literature, several critical research gaps motivate this investigation. First, majority of the existing research focuses on historical change detection rather than forecasting future morphological states. Second, although deep learning has been applied to water body segmentation and streamflow prediction, its application to multi-year spatial river morphology forecasting remains limited. Third, no systematic cross-architecture comparison for this specific forecasting task has been published. Fourth, physical geomorphological and socioeconomic perspectives are rarely integrated in unified predictive frameworks for operational disaster management.

To address these gaps, this study introduces SpaTempNet, an integrated spatiotemporal forecasting pipeline that: (1) employs sequence learning considering the continuous nature of river formation as a spatiotemporal process, using multiple years' windows to learn decadal changes; (2) performs a comparison of five architectures covering CNN+LSTM architectures and transformers under identical experimental settings; (3) extends the forecast horizon to 15 years for infrastructural planning and disaster preparedness; (4) incorporates an attention mechanism for automatic detection of morphologically dynamic regions; and (5) provides a base architecture offering future scope for inclusion of socioeconomic risk analysis for practical forecasting purposes.

II. Related Works

Riverbank erosion monitoring has evolved significantly with satellite remote sensing technology. Previously, researchers used to rely on aerial photography and manual channel boundary digitization. After five decades of continuous observation, the Landsat mission series revolutionized the field by enabling systematic, large-area, long-term monitoring at unprecedented scales.

A. Optical Satellite-Based Monitoring

Binh et al. [9] developed the Sentinel-based River Bank Erosion Detection (SRBED) framework for automated riverbank extraction from Landsat time-series. It demonstrated great potential for large-scale morphological change detection. Langhorst et al. [10] extended this globally. They constructed algorithms detecting lateral bank movement across diverse river systems and identified erosion hotspots at continental scales. These contributions established critical baseline erosion rates, which revealed extraordinary severity of riverbank retreat in South and Southeast Asia. Langat et al. [11] and Islam and Haque [12] demonstrated multi-temporal remote sensing with GIS for quantifying channel migration trends in Bangladesh. Raju et al. [13] analyzed erosion-accretion dynamics, and highlighted heterogeneous distribution of geomorphically active zones along river reaches where erosion alternates spatially with stable and accretionary zones.

B. Synthetic Aperture Radar Applications

Optical remote sensing in tropical environments suffers from persistent cloud cover, particularly during monsoon seasons when erosion peaks. Synthetic Aperture Radar (SAR) provides cloud-independent all-weather observation. Freihardt et al. [2] used Sentinel-1 SAR time series in Google Earth Engine to map riverbank erosion in Bangladesh. They documented rates exceeding 100 m/year and validated SAR water masking using VV backscatter threshold of -16 dB. DeVries et al. [23] demonstrated that combining Sentinel-1 and Landsat in Google Earth Engine enables rapid flood mapping scalable to continental extents. Sazib et al. [26] extended SAR-based mapping across the Ganga-Brahmaputra basin, and linked inundation extents with agricultural exposure and population impact assessments.

C. Regional Case Studies

Several studies focused on specific river systems. Tha et al. [14] evaluated riverbank erosion hotspots along Cambodia's Mekong River. They integrated physical erosion analysis with vulnerability assessments and connected geophysical processes with socioeconomic consequences. Saha [15] conducted multi-sensor analysis at the Padma-Jamuna confluence. They documented complex erosion, deposition, and island formation patterns from junction dynamics. Haque [16] studied multi-decadal Jamuna River changes and quantified agricultural land losses and settlement displacement. Ritu et al. [17] modeled Padma River bank shifting, however, 10-year intervals between acquisitions limited prediction accuracy and seasonal sensitivity. Muzahid et al. [18] quantified erosion and accretion along the Gorai River using multi-temporal Landsat classification.

D. Advanced Segmentation and Risk Assessment

Ren [19] developed large-scale erosion risk assessment, combined remote sensing with topographic, hydrological, and land-cover variables. A fundamental limitation shared across these studies is the absence of a multi-year predictive modeling grounded in spatiotemporal deep learning.

E. Socioeconomic Dimensions

Human dimensions of riverbank erosion provide strong motivation for improved predictive tools. Islam et al. [4] documented displacement patterns and economic losses. They showed erosion-induced migration drives rapid urbanization and informal settlement growth. Das et al. [5] documented psychological trauma from displacement and livelihood reconstruction impairment. They noted that the effects include not only economic losses but also damage to social ties and community strength. Alam et al. [3] and Sarkar et al. [6]

examined livelihood vulnerability among riparian households along the Padma and in northern Bangladesh. Their research revealed that poorest households face higher rate of exposure due to limited adaptive capacity and land-based livelihood dependence.

While qualitative vulnerability assessments help us see the human side of erosion, quantitative forecasting tools are necessary to give early warnings that people can act on. Some recent progress in deep learning and remote sensing now offers great chances to create these predictive systems. Below, we look at tech advances in water body segmentation, cloud-based processing, and spatiotemporal deep learning that form the base for today's river shape forecasting systems.

F. Deep Learning for Water Body Segmentation

Recent deep learning advances have replaced traditional threshold-based water extraction methods. Yuan et al. [20] demonstrated CNNs achieve superior accuracy over NDWI/MNDWI baselines across scenes with mixed land cover, shadows, and variable turbidity. Li et al. [21] proposed attention-enhanced multi-scale residual U-Net for water segmentation from Sentinel-2 imagery, Which showed attention mechanisms effectively discriminate water from spectrally similar classes like shadows and saturated soils. Ghosh et al. [22] advanced multi-scale fusion within U-Net architectures, They achieved enhanced detection and improved edge definition for rivers of varying dimensions. These architectures process individual images without temporal awareness and cannot forecast morphological change.

G. Cloud-Based Processing Infrastructure

Google Earth Engine (GEE) has made large-scale satellite processing accessible to all through operations on multi-decadal archives without local download. DeVries et al. [23] established platform suitability for operational flood monitoring. Clement et al. [24] applied GEE to ephemeral river flooding. Tripathy and Mallick [25] extended it to large-scale Mekong River monitoring, demonstrating continental scalability.

H. Recurrent and Hybrid Networks for Hydrological Forecasting

Fan et al. [7] applied LSTM to forecast river discharge and demonstrated that temporal relationships in hydrological data can be learned to generate superior predictions with autoregressive models. Le et al. [27] and Xiang et al. [28] validated hybrid CNN-LSTM frameworks for river flow prediction. Their research showed complementary gains from combining spatial feature extraction with temporal memory. Horvath et al. [29] highlighted importance of variable selection and temporal lag configuration for LSTM water level prediction on the Tisza River. Li et al. [21] demonstrated the value of attention mechanisms in CNN-LSTM models for streamflow forecasting across the Qinghai-Tibet Plateau.

I. ConvLSTM and Spatiotemporal Architectures

ConvLSTM architectures replace matrix multiplications in recurrent gates with convolutions which preserve spatial locality during temporal processing. Dehghani et al. [8] provided systematic comparative evaluation of LSTM, CNN, and ConvLSTM for spatiotemporal forecasting. They found ConvLSTM superior for tasks with strong spatial dependence without increasing computational cost. Lemenkova [1] applied deep learning, including convolutional-recurrent architectures, to track flood dynamics in the Ganges Delta. Their work demonstrated applicability to water extent data in the same geographical context. For transformer-based video understanding, Liu et al. [30] proposed Video Swin Transformer with hierarchical window-based self-attention. On the other hand Dosovitskiy et al. [31] established the ViT framework and Arnab et al. [32] extended it to video (ViViT). These researches provided architectural foundations for spatiotemporal transformer models evaluated here.

III. Study Area and Dataset Construction

A. Study Area: Padma River Mid-Course Reach

The designated study area (Fig. 1) covers a long mid-course section of the Padma River. It includes the active river channel, nearby floodplain zones, and the large islands and sandbar formations found in the middle of the river. The spatial region of interest was strategically chosen so as to incorporate effectively all possible patterns of morphological evolution and channel migration behavior within an analytically feasible spatial domain fit for training deep learning models intensively.

The Padma River forms through confluence of the Ganges (locally termed Padma upstream of junction) and Jamuna Rivers at Goalundo Ghat and flows approximately 120 km in southeastward direction before ultimately joining the Meghna River to discharge into the Bay of Bengal through multiple distributary channels. The river annually transports multiple millions of tonnes of Himalayan-derived suspended and bed-load sediment and exhibits characteristically highly dynamic morphodynamic behaviour including intense lateral channel migration, extreme bank erosion rates frequently exceeding 500 m during individual monsoon seasons, and regular mid-channel bar development, modification, and dissolution cycles. The discharge regime exhibits pronounced monsoonal characteristics: peak flow occurs during July–September monsoon season, during which the river maintains a complex braided planform geometry with rapidly changing channel configurations and extensive mid-channel island features. During dry season months (January–April), flow contracts substantially to form much narrower single-thread or limited multi-thread channel configurations with extensive exposed bar surfaces. The live river corridor provides high riparian population density on both bankside, with the river providing various socioeconomic roles including transportation for people and goods, irrigation waters for agriculture, artisan fishery for the riparian population, and clean water for domestic use. However, in spite of the above vital socioeconomic role, very high rates of erosion have resulted in loss of productive agricultural lands, forced displacement of entire households from their ancestral homelands, and complete destruction of vital infrastructure such as roads, bridges, ferry landing points, educational facilities, and health services facilities [3].

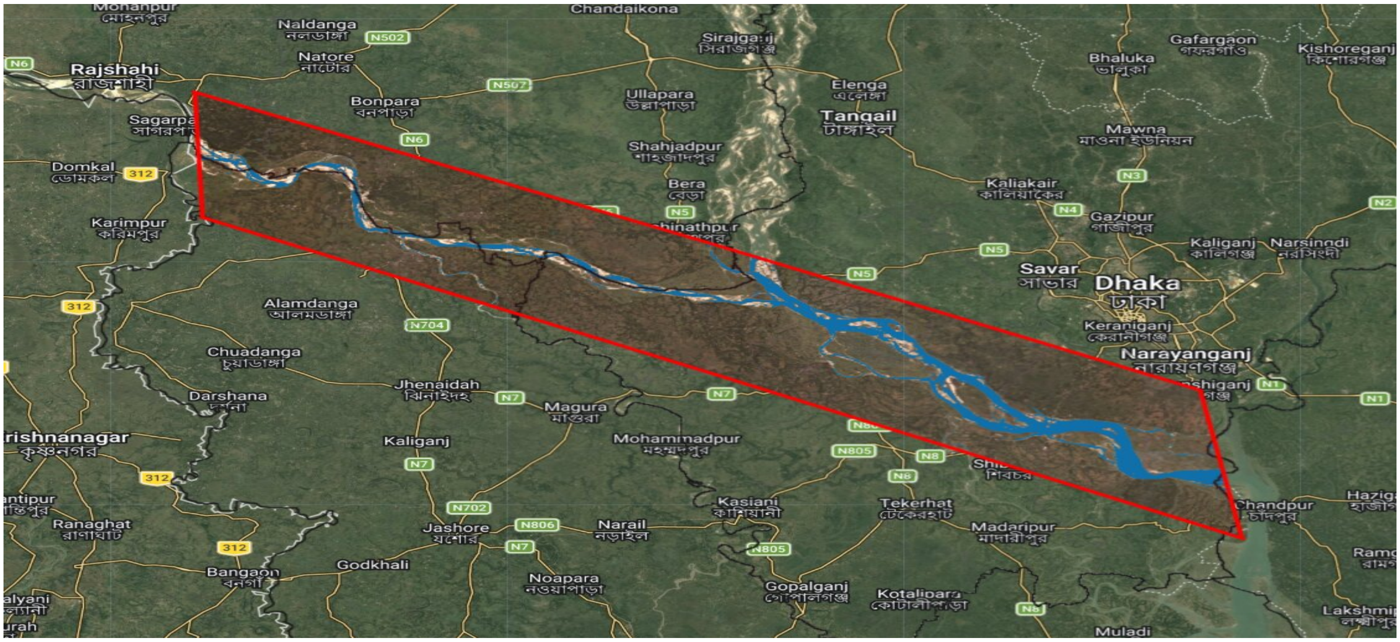

Fig. 1. Study area delineation along the mid-course section of the Padma River, Bangladesh.

B. Satellite Data Acquisition Protocol

Satellite imagery acquisition was strategically organized into two temporally distinct operational phases based on systematic sensor availability patterns and mission operational timelines.

1) Phase I: Landsat Mission Data (1987–2014).: Multispectral optical imagery was sourced exclusively from United States Geological Survey (USGS) Landsat Collection 2 Level-2 atmospherically-corrected surface reflectance product archives accessed through Google Earth Engine Application Programming Interface (API). Three successive sensor platforms were employed sequentially following their respective operational timelines: Landsat 5 Thematic Mapper (TM) sensor (GEE collection identifier: `LANDSAT/LT05/C02/T1_L2`) operational prior to 1999; Landsat 7 Enhanced Thematic Mapper Plus (ETM+) sensor (`LANDSAT/LE07/C02/T1_L2`) operational from 1999 onward; and Landsat 8 Operational Land Imager (OLI) sensor (`LANDSAT/LC08/C02/T1_L2`) operational from 2013 onward. All three sensors provide systematically consistent 60 m spatial resolution imagery (resampled from native 60 m for consistency with SWIR bands) with 16-day temporal revisit frequency. Scene-level cloud filtering applied maximum permissible cloud fraction threshold of 50% via `CLOUD_COVER` image-level metadata attribute. Pixel-level cloud and cloud-shadow masking protocols utilized `QA_PIXEL` quality assessment band via bitwise logical operations, specifically targeting bit 3 (cloud shadow detection) and bit 4 (cloud detection) binary flags. Landsat Collection 2 Level-2 products provide surface reflectance as scaled integers; conversion to physical surface reflectance values employed standardized scaling relationship:

$$\rho_\lambda = 0.0000275 \times \mathrm{DN}_\lambda - 0.2 \tag{1}$$

where, ρ_λ represents surface reflectance for spectral band λ and DN_λ represents digital number value. Spectral band selections for MNDWI computation varied systematically by sensor platform following band designation conventions: Band 2 (green, 520–600 nm) and Band 5 (SWIR1, 1550–1750 nm) for Landsat 5 and Landsat 7 platforms; Band 3 (green) and Band 6 (SWIR1) for Landsat 8 platform. Temporal composite imagery was systematically generated for three distinct temporal aggregation schemes: yearly (single annual composite per calendar year), quarterly (four three-month period composites per calendar year: Q1 January–March, Q2 April–June, Q3 July–September, Q4 October–December), and bi-monthly (six two-month period composites per calendar year: P1 January–February, P2 March–April, P3 May–June, P4 July–August, P5 September–October, P6 November–December).

2) Phase II: Sentinel Mission Data (2015–2025).: Sentinel-2 MultiSpectral Instrument (MSI) optical data were systematically accessed via Google Earth Engine collection identifier `COPERNICUS/S2_SR_HARMONIZED`, representing harmonized Level-2A bottom-of-atmosphere reflectance products. Scene-level filtering applied maximum cloud cover threshold of 50% based on `CLOUDY_PIXEL_PERCENTAGE` metadata attribute, with scenes subsequently sorted in ascending order of cloud fraction to prioritize least contaminated observations. To prevent computational timeout errors during large-scale export operations, maximum number of Sentinel-2 scenes utilized per temporal composite was algorithmically capped at 25 images. Pixel-level cloud masking protocols utilized QA60 quality assessment band targeting bit 10 (opaque cloud detection) and bit 11 (cirrus cloud detection) binary flags. Sentinel-2 MSI products provide surface reflectance as physical values requiring no additional radiometric scaling transformations.

To systematically address persistent data gap patterns arising from sustained cloud cover contamination particularly pronounced during monsoon seasons, Sentinel-1 Synthetic Aper-

TABLE I. SYSTEMATIC SUMMARY OF IMAGE AVAILABILITY AND QUALITY ASSESSMENT ACROSS MULTIPLE TEMPORAL RESOLUTION SCHEMES.

Temporal Resolution	Usable Images	Missing Images	Partial/ Low-Quality Images
Monthly	284	50	122
Bi-monthly	156	16	56
Quarterly	125	3	24
Yearly	38	0	0

ture Radar (SAR) C-band data were strategically incorporated as supplementary fallback data source activated when Sentinel-2 optical imagery proved unavailable. Sentinel-1 data were systematically accessed from Google Earth Engine collection `COPERNICUS/S1_GRD` representing Ground Range Detected (GRD) products acquired in Interferometric Wide (IW) swath mode with VV and VH dual polarization configuration in descending orbital pass geometry. Scene selection was sorted by acquisition date in descending temporal order and algorithmically limited to maximum of 20 images per temporal composite to manage computational processing load. The SAR-based fallback protocol was activated exclusively when zero Sentinel-2 scenes successfully passed cloud filtering criteria for specified temporal window. In this scenario, water mask generation involved using the backscatter coefficient of VV polarization, using an empirical fixed threshold of -16 dB, as per the protocol [2].

Systematically developed water masks were saved as georeferenced GeoTIFF files, having a geographic coordinate system EPSG:4326 (datum WGS84). The spatial resolution of 60 m was always used as a standard for all platforms, while falling back on bounding box size in the exceptional case of computation resource exhaustion.

C. Image Availability and Quality Assessment

Table I provides systematic summary of image availability characteristics and quality assessment metrics evaluated across four distinct temporal resolution aggregation schemes. The yearly dataset achieved complete temporal coverage (38 observations spanning 1988–2025, zero missing observations), thereby providing most stable and reliable foundation for long-term temporal trend analysis applications. The monthly resolution dataset exhibited most severe data loss patterns, with 50 completely missing temporal observations and 122 partial or low-quality degraded observations, thereby precluding its systematic utilization for primary modelling experiments. Bi-monthly and quarterly datasets demonstrated progressively improving completeness characteristics, with quarterly temporal resolution providing optimal balance between temporal granularity and systematic data reliability (only 3 completely missing observations across 38-year period).

Subsequent comprehensive manual inspection procedures complemented automated filtering protocols to systematically identify partially missing images exhibiting localized spatial gaps or degraded regions not adequately captured by cloud-fraction threshold criteria alone (Fig. 2). These quality impairment patterns proved most prevalent within bi-monthly dataset during July–August temporal periods (peak monsoon season) when persistent cloud cover systematically impedes optical observation capabilities.

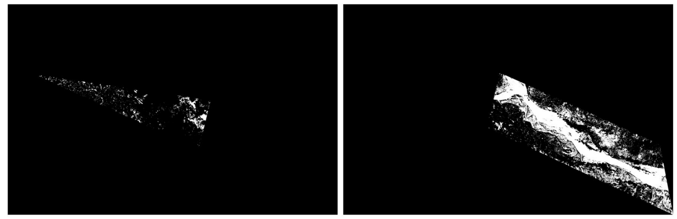

Fig. 2. Partially missing samples

D. Water Body Detection via Modified Normalized Difference Water Index

Water body extraction procedures employed Modified Normalised Difference Water Index (MNDWI) spectral index calculations [2], [14], which systematically exploit contrasting spectral reflectance behaviour characteristics of water surfaces (characteristically high reflectance in green visible wavelengths, strong absorption in shortwave infrared wavelengths) relative to terrestrial surfaces including soils, vegetation canopies, and built-up urban areas (producing characteristically low or negative MNDWI values). The MNDWI spectral index is formally defined:

$$\mathrm{MNDWI} = \frac{\rho_{\mathrm{Green}} - \rho_{\mathrm{SWIR}}}{\rho_{\mathrm{Green}} + \rho_{\mathrm{SWIR}}} \tag{2}$$

where, ρ_{Green} and ρ_{SWIR} represent surface reflectance values in green visible and shortwave infrared spectral bands, respectively. A fixed threshold value of zero was systematically applied to generate binary water/non-water classification masks from continuous MNDWI index values.

Following initial thresholding operations, connected component analysis algorithms were systematically applied to identify spatially contiguous water body regions. Only three largest spatially connected components per image were systematically retained, thereby eliminating isolated water pixels corresponding to small ponds, minor drainage channels, and ephemeral water features unrelated to primary Padma River channel. This connected component filtering substantially improves signal quality by removing spurious detections. All processed images were subsequently standardized via bilinear resampling to consistent 256×256 pixel spatial dimensions suitable for deep learning model input tensor specifications.

E. Gap-Filling Framework and Performance Benchmark

In order to effectively deal with natural patterns that will arise due to cloud contamination and instrument failure, a rigorous performance evaluation of gap filling methods was performed on a set of yearly water mask temporal series (a total of 38 observations from 1988 to 2025). For the purpose of simulating practical missing data problems, a systematic withholding process was used to remove 20% of the total number of observations (i.e., seven specific years: 2002, 2005, 2010, 2015, 2019, 2020, 2024).

There were seven unique strategies used to fill the gaps that underwent comparative analysis:

1) Temporal mean: The reference baseline strategy based on the use of temporal means to estimate the masks that need to be reconstructed, offering naive but effective and efficient spatial reconstructions of those masks.
2) Temporal median: The alternative reference baseline strategy using the median approach with enhanced robustness against outlier observation.
3) Linear temporal interpolation: Per-pixel linear interpolation computed between nearest available temporal neighbouring observations, yielding smooth temporal transition characteristics but exhibiting limited structural preservation capabilities for complex spatial features.
4) Cubic spline temporal interpolation: Per-pixel cubic spline interpolation applied across temporal dimension, enabling smooth nonlinear pixel trajectory modelling with continuous first and second derivatives providing enhanced temporal smoothness relative to linear interpolation.
5) Signed Distance Field (SDF) interpolation: Geometrically motivated approach wherein binary water masks are first transformed into signed distance field (SDF) representations defined as:
$$\mathrm{SDF}(\mathbf{x}) = d(\mathbf{x}, W^c) - d(\mathbf{x}, W) \tag{3}$$
where,
$$d(\mathbf{x}, A) = \min_{\mathbf{y}\in A} \|\mathbf{x} - \mathbf{y}\| \tag{4}$$
W denotes the water region, W^c represents its complement, $d(\mathbf{x}, A)$ is the Euclidean distance from location $\mathbf{x}$ to the set A, and $\|\cdot\|$ denotes the Euclidean norm. Positive SDF values correspond to pixels inside the water region, whereas negative values correspond to pixels outside the water region. Temporal interpolation is performed in the continuous SDF domain, followed by thresholding at the zero level set to recover binary water masks. This approach preserves boundary geometry and topological consistency more effectively than direct interpolation of binary masks.
6) Inverse-distance weighted temporal fusion: Missing mask reconstructions are computed through temporally weighted averaging of all available yearly observations, with weights inversely proportional to their temporal distance from the target year:
$$M_t = \frac{\sum_{i\in\mathcal{A}} w_i M_i}{\sum_{i\in\mathcal{A}} w_i}, \qquad w_i = \frac{1}{|t-i|} \tag{5}$$
where, M_t denotes the reconstructed mask at target year t, M_i represents an available observation at year i, and $\mathcal{A}$ denotes the set of all available observations. Consequently, observations temporally closer to the target year exert greater influence on the reconstruction than more distant observations.
7) Bidirectional Convolutional LSTM (BiConvLSTM): Deep learning-based approach wherein Bidirectional Convolutional LSTM architecture is systematically trained on 128×128 pixel spatiotemporal patches extracted randomly from full temporal series. Training protocol incorporates 20% random temporal masking simulation to replicate missing observation patterns. The model architecture learns both forward and backward temporal dependency patterns and was optimized using binary cross-entropy loss over maximum 100 training epochs with early stopping (patience = 15). Adam optimizer ($\eta = 10^{-3}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$) with batch size 16 was employed.

TABLE II. Comprehensive Performance Comparison of Seven Gap-Filling Methods Evaluated on Seven Held-Out Yearly Observations.

Method	IoU	Dice	Precision	Recall
Mean Composite	0.6573	0.7926	0.8535	0.7403
Median Composite	0.6573	0.7926	0.8535	0.7403
Linear Interp.	0.6548	0.7910	0.8913	0.7193
Spline Interp.	0.6733	0.8042	0.8082	0.8012
SDF	0.7290	0.8429	0.8741	0.8161
Weighted Temporal	0.6724	0.8035	0.8306	0.7790
BiConvLSTM	**0.7366**	**0.8477**	0.8457	**0.8511**

Quantitative performance results are comprehensively summarized in Table II. BiConvLSTM achieved best overall performance across all evaluation metrics (IoU = 0.7366, Dice = 0.8477, Recall = 0.8511), conclusively confirming its superior capability to capture extraordinarily complex spatiotemporal dynamic patterns. Among classical non-learning methods, SDF interpolation performed strongest (IoU = 0.7290, Dice = 0.8429), closely approaching deep learning performance levels while maintaining excellent structural consistency of water body boundary characteristics. Standard mean/median temporal composites and linear interpolation exhibited comparable baseline performance (IoU ≈ 0.655), while spline interpolation and weighted temporal fusion strategies offered modest incremental improvements.

IV. Methodology

This section of the study offers an in-depth technical explanation of all the methodological elements, such as sequence generation, temporal splitting, model architecture description with mathematical formulae, loss functions, metrics to evaluate the model, training parameters, and risk zone computation.

A. Supervised Sequence Generation Strategy

To formally formulate temporal morphological evolution process as supervised learning task, preprocessed binary water masks were represented as sequences of structured input-output pairs applying the technique of sliding temporal window. For every considered sliding window size $L \in \{4, 5, 6\}$ (with respect to yearly and quarterly temporal scales) or $L \in \{6, 9\}$ (with respect to bi-monthly temporal scale), precisely L sequential observations constitute input features tensor while next chronological time-step represents prediction target:

$$\mathbf{X}_i = (X_i, X_{i+1}, \ldots, X_{i+L-1}), \quad y_i = X_{i+L} \tag{6}$$

where, $X_t \in \{0, 1\}^{H \times W}$ represents binary water mask at time t with spatial dimensions $H \times W = 256 \times 256$ pixels. Temporal stride of one year (or one bi-monthly period for higher-frequency data) generates systematically overlapping sequences, maximizing utilization of available but limited temporal data. This multi-sequence evaluation protocol enables systematic identification of optimal temporal context window length for each distinct architecture and temporal resolution combination.

B. Rigorous Temporal Splitting Strategy

A meticulously rigorous temporal split strategy prevents any data leakage between training and evaluation datasets. Two experimental evaluation setups are formally defined.

1) Setup 1 (Long-range forecasting): Training incorporates all available data up to and including calendar year 2015; testing evaluates forecast performance on 2016–2025 inclusive (providing 10-year forecasting horizon for yearly data, or equivalent temporal span for quarterly/bi-monthly aggregations).

2) Setup 2 (Medium-range forecasting): Training incorporates all available data up to and including calendar year 2020; testing evaluates forecast performance on 2021–2025 inclusive (providing 5-year forecasting horizon for yearly data, or equivalent temporal span for quarterly/bi-monthly aggregations).

For quarterly temporal resolution data, temporal cutoffs correspond to Q4-2015 and Q4-2020, respectively. For bi-monthly resolution, cutoffs correspond to P6-2015 and P6-2020 (November–December 2015 and 2020). Within each setup configuration, final 15% of training sequences (ordered chronologically) are systematically allocated to validation dataset for hyperparameter tuning and early stopping criteria. Critically, absolutely zero temporal overlap exists between training, validation, and testing datasets, ensuring evaluation integrity.

C. Model Architecture Specifications

Five distinct spatiotemporal deep learning architectures underwent systematic comparative evaluation, selected to comprehensively span the architectural spectrum from recurrent convolutional hybrids to pure attention-based transformer models.

1) Standalone ConvLSTM architecture: ConvLSTM framework [8] primarily involves the replacement of traditional matrix multiplications in LSTMs with spatial convolutions. This allows inherent joint processing of spatial and temporal information without any need for a dedicated encoder module. The overall architecture includes:

a) Initial spatial feature extraction: Two successive spatial Conv2D layers with 32 and 64 convolutional filters (kernel size 3×3, ReLU activation, same padding) extract initial low-level spatial features.

b) Spatial downsampling: Batch normalization followed by max-pooling (2×2 window, stride 2) and spatial dropout (rate = 0.2) applied after each Conv2D layer.

c) Spatiotemporal processing: Two stacked ConvLSTM2D layers with 128 and 64 filters respectively (kernel size 3×3). The first ConvLSTM2D layer processes full input sequence with `return_sequences=True`; the second layer aggregates temporal information with `return_sequences=False`, outputting single feature map.

d) Upsampling decoder: Symmetric decoder comprising two upsampling blocks (UpSampling2D with factor 2 followed by Conv2D with 64 and 32 filters), batch normalization, and final 1×1 Conv2D with sigmoid activation producing prediction probability map.

2) U-Net with LSTM architecture: The U-Net+LSTM hybrid model [27], [28] processes input sequence through TimeDistributed encoder extracting multi-scale spatial features independently for each temporal frame. Complete architecture specification:

a) Timedistributed encoder: Three successive encoder blocks each comprising: Conv2D layer (filters: 64, 128, 256 for blocks 1–3; kernel 3×3, ReLU), batch normalization, second Conv2D with identical specifications, max-pooling (2×2), spatial dropout (rate = 0.2). TimeDistributed wrapper ensures identical processing applied to each input frame independently.

b) Bottleneck temporal fusion: LSTM module (128 units) at encoder bottleneck models temporal dependencies across sequence, producing single temporally-fused feature volume.

c) Symmetric decoder with skip connections: Three decoder blocks each comprising: UpSampling2D (factor 2), concatenation with corresponding encoder skip connection feature map, two Conv2D layers (filters decreasing: 256, 128, 64), batch normalization. Final 1×1 Conv2D with sigmoid activation produces prediction.

3) Attention U-Net with ConvLSTM architecture: The Attention U-Net+ConvLSTM model extends U-Net+LSTM through incorporating sophisticated attention gate mechanisms in each skip connection pathway [21], [22]. For encoder feature map $\mathbf{x}^l \in \mathbb{R}^{H_l \times W_l \times C_l}$ at encoder level l and gating signal $\mathbf{g} \in \mathbb{R}^{H_l \times W_l \times C_g}$ derived from corresponding decoder level, attention coefficient computation proceeds via:

$$\mathbf{q}^l = W_x * \mathbf{x}^l + W_g * \mathbf{g} + b_g \tag{7}$$

$$\boldsymbol{\alpha}^l = \sigma\big(W_\psi * \mathrm{ReLU}(\mathbf{q}^l) + b_\psi\big) \tag{8}$$

where, $*$ denotes 1×1 convolution operation, W_x, W_g, W_ψ represent learnable convolutional kernel parameters, b_g, b_ψ are bias terms, and σ denotes sigmoid activation function producing spatial attention coefficients $\boldsymbol{\alpha}^l \in [0, 1]^{H_l \times W_l}$. The skip connection output becomes element-wise product:

$$\tilde{\mathbf{x}}^l = \boldsymbol{\alpha}^l \odot \mathbf{x}^l \tag{9}$$

thereby suppressing background regions (low α values) and amplifying morphologically active zones including channel margins and active erosion fronts (high α values). Critically, the bottleneck employs ConvLSTM2D rather than standard LSTM, preserving spatial locality during temporal fusion operations.

4) Swin Spatio-Temporal Transformer architecture: The Swin Spatio-Temporal Transformer [30] applies shared hierarchical Swin Transformer encoder to each input frame independently. The complete architecture specification:

a) Shared swin encoder: Each stage applies window-based multi-head self-attention with shifted window partitioning strategy (W-MSA and SW-MSA) followed by patch merging for progressive spatial downsampling. Three encoder stages produce feature maps at spatial resolutions $H/4 \times W/4$, $H/8 \times W/8$, and $H/16 \times W/16$ with embedding dimensions 64, 128, and 256, respectively. Self-attention within each window $\mathcal{W}$ of size $M \times M$ is computed as:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{SoftMax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}} + \mathbf{B}\right)\mathbf{V} \quad (10)$$

where, $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ represent query, key, value projections, d is head dimension, and $\mathbf{B}$ encodes relative position biases. Shifted windowing across successive layers enables cross-window connections.

b) Temporal fusion module: Features at finest encoder resolution across all T input frames are stacked forming spatiotemporal tensor $\mathbf{F} \in \mathbb{R}^{T \times (H/4) \times (W/4) \times 64}$. Temporal fusion comprises:

- 1D temporal convolution (kernel size 3, GELU activation) applied across temporal dimension at each spatial location.
- Two cross-temporal attention layers (4 heads each) enabling each spatial location to attend across all temporal frames.
- Softmax-based temporal aggregation producing single fused feature map $\bar{\mathbf{F}} \in \mathbb{R}^{(H/4) \times (W/4) \times 64}$.

c) U-Net style decoder: Symmetric decoder with skip connections from encoder stages 1 and 2 progressively upsamples to full resolution.

5) Vision Transformer-based spatio-temporal model: The ViT-based model [31], [32] applies shared Vision Transformer encoder to each frame independently. Complete specification:

a) Patch embedding: Each frame is partitioned into non-overlapping 16×16 patches yielding $(256/16)^2 = 256$ patch tokens. Linear projection maps each flattened patch to 128-dimensional embedding space:

$$\mathbf{z}_p = \mathbf{E} \cdot \text{vec}(\mathbf{x}_p) + \mathbf{e}_{\text{pos}} \quad (11)$$

where, $\mathbf{E} \in \mathbb{R}^{128 \times (16 \times 16)}$ represents learnable projection matrix, $\mathbf{e}_{\text{pos}}$ encodes spatial position.

b) Spatial transformer blocks: Two successive transformer blocks (multi-head self-attention with 4 heads, MLP with hidden dimension 256, LayerNorm, residual connections, dropout 0.1) process patch token sequence.

c) Temporal aggregation: Global average pooling aggregates all patch tokens yielding per-frame feature vector $\mathbf{f}_t \in \mathbb{R}^{128}$. Frame features are stacked, augmented with learned temporal positional embeddings, and processed by two temporal transformer layers (4 heads, MLP dimension 256) modeling inter-frame dependencies.

d) Convolutional decoder: Aggregated temporal feature is spatially projected to $16 \times 16 \times 128$ feature map, then decoded through four transposed convolution blocks progressively upsampling to full resolution.

D. Loss Function Specifications

All models underwent training optimization using combined Dice and Focal loss formulation:

$$\mathcal{L}_{\text{Total}} = \lambda_1 \mathcal{L}_{\text{Dice}} + \lambda_2 \mathcal{L}_{\text{Focal}} \quad (12)$$

where, weighting coefficients $\lambda_1 = \lambda_2 = 0.5$ provide equal contribution.

1) Dice loss: directly optimizes Dice overlap metric and proves particularly effective for class-imbalanced segmentation where water pixels constitute spatial minority:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_i p_i g_i + \varepsilon}{\sum_i p_i + \sum_i g_i + \varepsilon} \quad (13)$$

where, p_i represents predicted probability for pixel i, $g_i \in \{0, 1\}$ represents ground truth label, and $\varepsilon = 1$ provides numerical stability smoothing.

2) Focal loss: systematically down-weights easily-classified background examples to concentrate gradient updates on challenging boundary pixels [33]:

$$\mathcal{L}_{\text{Focal}} = -\frac{1}{N}\sum_{i=1}^{N} \alpha_i (1 - p_{t_i})^{\gamma} \log(p_{t_i}) \quad (14)$$

where, $p_{t_i} = p_i$ if $g_i = 1$ else $1 - p_i$, class weighting $\alpha_i = 0.75$ for positive class, focusing parameter $\gamma = 2.0$, and N represents total pixel count. The loss functions work together by utilizing each other's strengths: Dice loss guarantees accuracy with regards to overall overlap, Focal loss supplies accurate supervision on the level of boundaries.

E. Comprehensive Evaluation Metrics

The model was evaluated with five complementary quantitative metrics using the binarization threshold value of 0.5:

Intersection over Union (IoU):

$$\text{IoU} = \frac{|A \cap B| + \varepsilon}{|A \cup B| + \varepsilon} = \frac{\text{TP} + \varepsilon}{\text{TP} + \text{FP} + \text{FN} + \varepsilon} \quad (15)$$

Dice Coefficient (F1-Score):

$$\text{Dice} = \frac{2|A \cap B| + \varepsilon}{|A| + |B| + \varepsilon} = \frac{2\,\text{TP} + \varepsilon}{2\,\text{TP} + \text{FP} + \text{FN} + \varepsilon} \quad (16)$$

Precision:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{17}$$

Recall (Sensitivity):

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{18}$$

where, TP, FP, FN represent true positives, false positives, and false negatives, respectively, $\varepsilon = 10^{-7}$ provides numerical stability.

Additionally, absolute area difference quantifies spatial calibration independently of overlap metrics:

$$\Delta A = A_{\text{pred}} - A_{\text{true}} = (N_{\text{pred}} - N_{\text{true}}) \times A_{\text{pixel}} \tag{19}$$

where, N represents pixel counts, $A_{\text{pixel}} = 0.0036$ km^2 at 60 m resolution. A persistence baseline (predicting last observed frame) provides reference lower bound.

F. Training Configuration and Hyperparameters

1) CNN-LSTM hybrid models (ConvLSTM, U-Net+LSTM, Attention U-Net+ConvLSTM): All three architectures employed identical training configuration for fair comparison. Optimization utilized Adam algorithm [38] with learning rate $\eta = 1 \times 10^{-4}$, momentum parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and gradient clipping via `clipnorm` = 1.0 preventing exploding gradient instability. Batch size of 4 sequences balanced GPU memory constraints against gradient estimate quality. Training proceeded for maximum 200 epochs with early stopping monitoring validation loss (patience = 20 epochs, restore best weights). ReduceLROnPlateau callback dynamically adjusted learning rate upon validation loss plateau detection (patience = 7 epochs, reduction factor = 0.5, minimum $\eta_{\min} = 1 \times 10^{-7}$). Model checkpointing systematically saved best-performing checkpoint based on validation loss criterion.

2) Swin Spatio-Temporal Transformer: Reduced batch size of 2 sequences accommodated substantially, elevated GPU memory requirements (patch size = 4, embedding dimension = 64, window size = 8, hierarchical depths $[2, 2, 2]$, attention heads $[2, 4, 8]$, MLP ratio = 2.0, spatial dropout = 0.1). Temporal fusion module configured with 2 temporal transformer layers, 4 attention heads, MLP dimension 256, dropout 0.1. Decoder filter progression: $[128, 64, 32, 16]$. Remaining training hyperparameters identical to CNN-LSTM models.

3) ViT-based spatio-temporal model: Batch size 2 (patch size = 16, embedding dimension = 128, 2 spatial transformer layers, 2 temporal transformer layers, 4 attention heads each, MLP dimension 256, dropout 0.1). Remaining hyperparameters identical.

All model implementations used deterministic random seeds (NumPy: 42, TensorFlow: 42) ensuring repeatability and consistent data splitting for controlled comparisons.

G. Computational Framework for Risk Zone Mapping

After iterative autoregressive forecasting for the time span 2026-2045, post-processing framework is employed to define erosion and accretion risk zones using pixel-resolution spatial data.

Erosion strictly defined as the land to water transition between consecutive years ($L_t = 0 \wedge L_{t+1} = 1$), while **accretion** is a reverse process (water to land). Frequency maps characterize persistence duration:

$$f_{\text{erosion}}(\mathbf{x}) = \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}(L_t(\mathbf{x}) = 0 \wedge L_{t+1}(\mathbf{x}) = 1) \tag{20}$$

$$f_{\text{accretion}}(\mathbf{x}) = \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{I}(L_t(\mathbf{x}) = 1 \wedge L_{t+1}(\mathbf{x}) = 0) \tag{21}$$

where, $\mathbf{x}$ denotes spatial pixel location, T represents total forecast years, $\mathbb{I}(\cdot)$ is indicator function.

A **hydrodynamic stability index** quantifies pixel-level classification uncertainty via water occurrence frequency:

$$\mathcal{I}(\mathbf{x}) = 1 - |2F_w(\mathbf{x}) - 1|, \quad F_w(\mathbf{x}) = \frac{1}{T} \sum_{t=1}^{T} L_t(\mathbf{x}) \tag{22}$$

where, $F_w \in [0, 1]$ represents fraction of forecast period classified as water. Values $\mathcal{I} \approx 1$ indicate persistent land–water oscillation (maximum instability); $\mathcal{I} \approx 0$ indicates stable land or water classification (maximum stability).

Composite risk representation integrates erosion frequency, accretion frequency, and instability index highlighting regions simultaneously dynamic and uncertain. Physical area estimates employ:

$$A_{\text{erosion/accretion}} = N_{\text{pixels}} \times A_{\text{pixel}} \tag{23}$$

where, $A_{\text{pixel}} = (60 \text{ m})^2 = 0.0036$ km^2 at established 60 m spatial resolution.

V. Spatiotemporal Statistical Analysis

This section presents comprehensive quantitative spatiotemporal statistical characterization of 38 years of Padma River morphodynamic behaviour (1988–2025) including long-term areal dynamics, seasonal variability patterns, erosion–accretion mass budget calculations, and channel centreline migration rate quantification.

A. Long-Term Areal Dynamics and Trend Analysis

The Padma River exhibited substantial inter-annual variability in water-covered areal extent throughout 38-year observation period. Annual water-covered area ranged from absolute minimum 540.4 km^2 (calendar year 1995, hydrologically dry year characterized by below-average monsoon precipitation) to absolute maximum 1,229.3 km^2 (calendar year 1999, catastrophic monsoon flood corresponding to major regional disaster event), with long-term temporal mean of 664.5 $\pm$ 112.8 km^2 (arithmetic mean $\pm$ one standard deviation). The exceptionally large channel expansion manifest in 1999 corresponds temporally to the catastrophic July–August 1998 monsoon flood event representing one of most severe recorded flood disasters in Bangladesh's documented history.

Formal non-parametric Mann–Kendall trend test [34], [35] indicated absence of statistically significant monotonic trend at $\alpha = 0.05$ significance level (Kendall's $\tau = -0.149$, two-tailed $p = 0.187$), and Theil–Sen robust linear estimator [36], [37] yielded weak negative slope magnitude of -1.41 km^2 yr^{-1}. The above complementary results of statistical analysis have effectively validated the fact that although there is a marked variation in the temporal pattern due to monsoon variations, the mean temporal pattern of river area does not exhibit any clear trend toward increase or decrease.

B. Seasonal Variability and Intra-Annual Dynamics

Magnitudes of seasonality were systematically analyzed by way of coefficient of variation (CV) computed for the corresponding time-series of bi-monthly water surface area magnitudes, revealing significant inter-annual variability (see Fig. 3). Magnitude of annual CV varied from a lowest value of 8.2% (year 2018, most hydrologically stable year with little seasonal change), to a highest of 64.1% (year 2007, most hydrologically unstable year), with an arithmetic mean value for the study interval of 34.3 $\pm$ 13.7%. Those calendar years that had high CV values, especially 2007, 2011, and 2020, were systematically related to extreme monsoon flooding episodes, characterized by a maximum water extent above 2,700 km^2. Years with lower CV values indicate higher stability of seasonal trajectory of water extent as well as comparatively smaller magnitude of variation of the flood-pulse.

Heat map representation of the quarterly data clearly reveals persistent and highly consistent hydrological seasonality pattern across all 38 years of monitoring; specifically, Q4 (October-December; post-monsoon recession season) invariably shows largest mean water extent, while Q2 (April-June; pre-monsoon low-water flow season) invariably records minimum extent values. This remarkable hydrological seasonality pattern clearly indicates the predominant influence of the South Asian monsoon meteorological system on hydrodynamics.

C. Erosion and Accretion Dynamics

Annual erosion and accretion rate magnitudes demonstrated exceptionally high inter-annual variability reflecting pronounced sensitivity to individual monsoon flood events (Fig. 4). The largest documented single-year erosion event occurred during 1998-1999 interval (649.1 km^2 land surface converted to water classification; net channel expansion 567.7 km^2), corresponding temporally to catastrophic July-August 1998 monsoon flood representing most extreme channel widening event recorded throughout entire 38-year study period. Conversely, the immediately following year (1999–2000 interval) recorded largest net accretion magnitude (-503.5 km^2), indicating substantial rapid channel contraction during post-flood geomorphic recovery and adjustment phase. This dramatic reversal pattern highlights pronounced recovery dynamics characterizing post-extreme-event channel behaviour.

TABLE III. DECADAL EROSION, ACCRETION, AND NET CHANNEL CHANGE SUMMARY.

Decadal Period	Erosion (km^2)	Accretion (km^2)	Net Change (km^2)
1988–1998	367.6	434.2	-66.7
1998–2008	382.3	355.9	$+26.3$
2008–2018	262.8	375.8	-113.0
2018–2025	264.6	264.3	$+0.3$
Total	**5170.2**	**5323.3**	$\mathbf{-153.0}$

Table III provides systematic quantitative summary of decadal erosion–accretion mass budgets.

Across complete 38-year study period, cumulative gross erosion reached 5,170.2 km^2 (representing multiple complete channel reworking cycles) while cumulative gross accretion totalled 5,323.3 km^2, yielding overall net accretion magnitude of 153.0 km^2 indicating slight long-term channel contraction tendency. Decadal analysis systematically identifies four temporally distinct morphodynamic regime phases: (1) 1988-1998 exhibited near-balanced erosion–accretion dynamics (-66.7 km^2 net accretion) characteristic of lateral channel migration without sustained widening; (2) 1998–2008 shifted toward slight net erosion tendency ($+26.3$ km^2) dominated by catastrophic 1998–1999 flood event and subsequent multi-year adjustment; (3) 2008–2018 manifested strongest net accretion (-113.0 km^2) suggesting significant channel contraction and partial geomorphic stabilization; (4) 2018–2025 approached near-equilibrium conditions with negligible net change ($+0.3$ km^2) indicating contemporary period of relative morphodynamic stability.

D. Channel Centreline Migration Rate Quantification

Channel centreline lateral migration rates exhibited pronounced temporal variability reflecting alternating periods of geomorphic instability and relative stabilization (Fig. 5). Arithmetic mean annual migration rate across full study period was 255.4 m yr^{-1}, with individual year magnitudes ranging from minimum 134.2 m yr^{-1} (calendar years 2001-2002 and 2022-2023, representing temporary stabilization phases) to catastrophic maximum 1,485.5 m yr^{-1} during 1998–1999 interval (corresponding to extreme flood-driven channel adjustment event). The median migration rate (216.3 m yr^{-1}) was substantially below arithmetic mean, conclusively confirming that small number of extreme flood years strongly inflated long-term average metrics while typical inter-annual behaviour exhibits more moderate migration magnitudes. The 1990s decade recorded highest average migration rates (mean 348.7 m yr^{-1}), definitively designating that decade as period of maximum geomorphic instability within 38-year observational record. Cumulative centreline displacement over complete study period exceeded 9 km total lateral adjustment, demonstrating sustained long-term lateral migration and continuous comprehensive reworking of active river corridor.

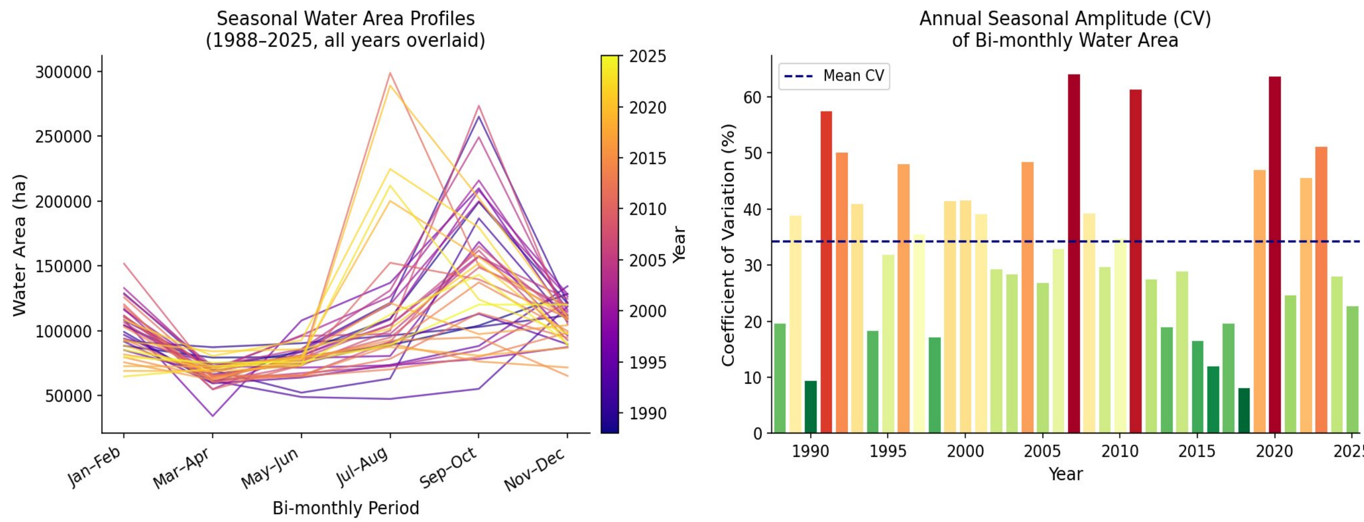

Fig. 3. Seasonal water area dynamics. Left: Bi-monthly water area profiles for all years. Right: Annual coefficient of variation (CV) of bi-monthly water area.

Fig. 4. Annual and decadal erosion and accretion of the channel. Top: annual gross erosion (red), gross accretion (blue), and net channel change (black line). Bottom: decadal grouped summary illustrating long-term morphodynamic phases.

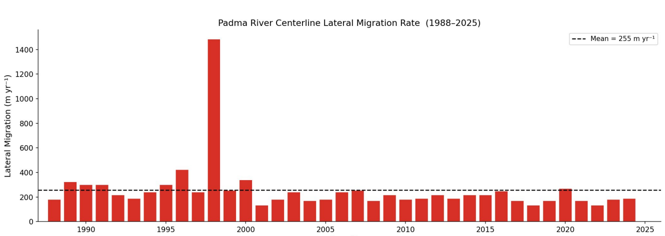

Fig. 5. Annual lateral migration rate of the centreline. Red bars indicate yearly migration rates, while the dashed line marks the long-term mean of 255.4 m yr^{-1}. The extreme peak in 1998–1999 corresponds to the catastrophic flood-driven channel adjustment.

Attention U-Net+ConvLSTM achieved highest IoU (0.7005) and Dice (0.8236) in Setup 1, establishing it as optimal architecture for yearly prediction tasks. U-Net+LSTM demonstrated competitive performance with slightly lower overlap metrics but superior area calibration. ConvLSTM and Swin Transformer exhibited moderate performance, while Vision Transformer significantly underperformed all alternative architectures (see Table IV).

TABLE IV. Performance Comparison of all Architectures Based on Yearly Data.

Model	Setup	IoU	Dice	P	R	$\|\Delta A\|$
ConvLSTM	S1	0.6218	0.7665	0.7212	0.8207	76.69
ConvLSTM	S2	0.6278	0.7711	0.7139	0.8432	103.26
U-Net+LSTM	S1	0.6964	0.8208	0.7876	0.8595	50.25
U-Net+LSTM	S2	0.6986	0.8223	0.7905	0.8592	48.61
Att.-U-Net+CLSTM	S1	**0.7005**	**0.8236**	0.7807	0.8747	66.54
Att.-U-Net+CLSTM	S2	0.6963	0.8205	0.7934	0.8523	**41.05**
ViT	S1	0.3613	0.5302	0.4352	0.6806	316.06
ViT	S2	0.3831	0.5526	0.5197	0.5907	78.69
Swin	S1	0.6080	0.7560	0.7364	0.7780	31.64
Swin	S2	0.5912	0.7431	0.6849	0.8122	96.23

1) Detailed year-wise analysis: Table V and Table VI present comprehensive year-by-year breakdown for best-performing Attention U-Net+ConvLSTM architecture.

Performance remained remarkably stable across 10-year horizon with IoU ranging 0.6569–0.7359 (standard deviation 0.0273). Higher Recall (mean 0.8747) relative to Precision (0.7807) indicates systematic tendency toward over-prediction, consistent with positive area differences in most years. The 5-year setup achieved lower area differences (mean 41.05 km^2 vs 66.54 km^2), confirming shorter horizons reduce temporal uncertainty.

TABLE V. DETAILED YEAR-WISE PREDICTION RESULTS: ATTENTION U-NET+CONVLSTM, 10-YEAR FORECASTING SETUP (2016–2025).

Year	IoU	Dice	Prec.	Rec.	ΔA (km^2)
2016	0.7357	0.8477	0.8655	0.8307	-24.96
2017	0.6975	0.8218	0.7396	0.9246	130.72
2018	0.6742	0.8054	0.7273	0.9023	122.18
2019	0.7051	0.8270	0.7858	0.8729	58.46
2020	0.7055	0.8273	0.7924	0.8654	51.24
2021	0.6569	0.7929	0.7706	0.8165	34.81
2022	0.7359	0.8479	0.8407	0.8552	10.84
2023	0.7081	0.8291	0.7618	0.9095	107.73
2024	0.7046	0.8267	0.7806	0.8786	70.94
2025	0.6814	0.8105	0.7430	0.8915	103.46
Mean	**0.7005**	**0.8236**	0.7807	0.8747	66.54
Std	0.0273	0.0179	0.0440	0.0325	56.23

TABLE VI. DETAILED YEAR-WISE PREDICTION RESULTS: ATTENTION U-NET+CONVLSTM, 5-YEAR FORECASTING SETUP (2021–2025).

Year	IoU	Dice	Prec.	Rec.	ΔA (km^2)
2021	0.6384	0.7793	0.7797	0.7788	-0.66
2022	0.7270	0.8419	0.8546	0.8296	-18.39
2023	0.7143	0.8333	0.7745	0.9018	91.31
2024	0.7000	0.8235	0.7950	0.8542	42.04
2025	0.7017	0.8247	0.7631	0.8972	90.98
Mean	0.6963	0.8205	0.7934	0.8523	41.05
Std	0.0341	0.0244	0.0334	0.0511	50.98

TABLE VII. PERFORMANCE COMPARISON OF ALL ARCHITECTURES BASED ON QUARTERLY DATA.

Model	Setup	IoU	Dice	P	R	$\|\Delta A\|$
ConvLSTM	S1	0.6114	0.7572	0.7433	0.7918	166.66
ConvLSTM	S2	0.6134	0.7586	0.7115	0.8348	159.91
U-Net+LSTM	S1	0.6908	0.8095	0.8162	0.8318	188.03
U-Net+LSTM	S2	**0.7008**	0.8191	0.8257	0.8431	163.76
Att.-U-Net+CLSTM	S1	**0.6956**	**0.8139**	0.8095	**0.8443**	184.80
Att.-U-Net+CLSTM	S2	0.6972	**0.8170**	0.8054	0.8511	**137.98**
ViViT	S1	0.3648	0.5316	0.4788	0.6347	340.08
ViViT	S2	0.3822	0.5496	0.4606	0.7281	447.09
Swin	S1	0.5865	0.7370	0.7012	0.7981	190.79
Swin	S2	0.5882	0.7392	0.7004	0.8024	160.45

E. Quarterly Prediction Results: Complete Architecture Comparison

Attention U-Net+ConvLSTM and U-Net+LSTM exhibited closely competitive performance for quarterly prediction, with Attention model achieving marginally superior IoU in Setup 1 (0.6956 vs 0.6908) (see Table VII).

1) Detailed quarterly results Setup 1: Table VIII presents complete quarter-by-quarter breakdown for Attention U-Net+ConvLSTM covering 2017 Q2–2025 Q4 test period.

TABLE VIII. DETAILED QUARTERLY PREDICTION RESULTS – SETUP 1 (ATTENTION U-NET+CONVLSTM, SEQLEN=8, TEST: 2017 Q2–2025 Q4).

Quarter	IoU	Dice	Prec.	Rec.	ΔA (km^2)
2017-Q2	0.9156	0.9559	0.9524	0.9595	5.31
2017-Q3	0.7556	0.8608	0.9543	0.7840	-171.18
2017-Q4	0.7778	0.8750	0.9127	0.8404	-73.25
2018-Q1	0.8315	0.9080	0.8833	0.9341	41.75
2018-Q2	0.9150	0.9556	0.9304	0.9823	38.33
2018-Q3	0.7056	0.8274	0.8341	0.8208	-12.15
2018-Q4	0.5594	0.7175	0.6733	0.7678	94.13
2019-Q1	0.8782	0.9351	0.8851	0.9911	71.73
2019-Q2	0.6949	0.8200	0.7197	0.9528	197.74
2019-Q3	0.5647	0.7218	0.9123	0.5971	-447.49
2019-Q4	0.6155	0.7620	0.6746	0.8754	241.01
2020-Q1	0.7440	0.8532	0.7610	0.9708	157.89
2020-Q2	0.9068	0.9511	0.9353	0.9675	20.12
2020-Q3	0.5061	0.6721	0.8699	0.5475	-565.15
2020-Q4	0.5661	0.7230	0.6572	0.8034	179.91
2021-Q1	0.7922	0.8840	0.8376	0.9360	72.49
2021-Q2	0.8128	0.8967	0.8156	0.9958	119.94
2021-Q3	0.4833	0.6517	0.8805	0.5173	-662.31
2021-Q4	0.5279	0.6910	0.5910	0.8316	320.34
2022-Q1	0.6335	0.7756	0.6908	0.8842	142.33
2022-Q2	0.6190	0.7647	0.9081	0.6604	-213.69
2022-Q3	0.5230	0.6868	0.8612	0.5711	-519.60
2022-Q4	0.6256	0.7697	0.7454	0.7956	63.38
2023-Q1	0.7641	0.8663	0.8414	0.8926	38.33
2023-Q2	0.7147	0.8336	0.7684	0.9109	107.41
2023-Q3	0.6263	0.7702	0.9154	0.6647	-306.30
2023-Q4	0.6325	0.7749	0.6465	0.9670	359.43
2024-Q1	0.6504	0.7882	0.6631	0.9713	215.20
2024-Q2	0.7352	0.8474	0.8906	0.8081	-52.76
2024-Q3	0.5501	0.7097	0.5527	0.9915	427.37
2024-Q4	0.6480	0.7864	0.7055	0.8882	174.97
2025-Q1	0.8124	0.8965	0.8238	0.9833	118.42
2025-Q2	0.9298	0.9636	0.9493	0.9783	18.22
2025-Q3	0.5993	0.7495	0.8262	0.6858	-170.04
2025-Q4	0.7280	0.8426	0.8655	0.8209	-48.20
Mean	**0.6956**	**0.8139**	0.8095	0.8443	-0.47
Std	0.1230	0.0851	0.1124	0.1275	244.68

Pronounced seasonal stratification is evident: Q2 (pre-monsoon) consistently achieves IoU > 0.90 in multiple years (2017, 2018, 2020, 2025), while Q3 (peak monsoon) exhibits lowest performance, frequently falling below 0.55. The signed mean area difference approaches zero (-0.47 km^2), indicating excellent long-term calibration despite substantial within-year fluctuations driven by seasonal effects.

2) Detailed quarterly results Setup 2: Table IX presents detailed results for shorter 5-year forecasting horizon (2022 Q4–2025 Q4).

Setup 2 results demonstrate consistent seasonal patterns with improved mean performance (IoU = 0.6972 vs. 0.6956) attributable to shorter forecasting horizon reducing cumulative error propagation.

F. Bi-Monthly Prediction Results: Complete Architecture Comparison

U-Net+LSTM emerged as best-performing architecture for bi-monthly prediction, achieving IoU = 0.7791 and Dice = 0.8701 in Setup 2. This architectural shift from

TABLE IX. DETAILED QUARTERLY PREDICTION RESULTS – SETUP 2 (ATTENTION U-NET+CONVLSTM, SEQLEN=10, TEST: 2022 Q4–2025 Q4)

Quarter	IoU	Dice	Prec.	Rec.	ΔA (km^2)
2022-Q4	0.6086	0.7567	0.7498	0.7637	17.46
2023-Q1	0.7485	0.8561	0.8325	0.8812	36.82
2023-Q2	0.7025	0.8253	0.7826	0.8729	66.80
2023-Q3	0.6164	0.7627	0.8820	0.6719	−266.44
2023-Q4	0.6406	0.7810	0.6768	0.9230	263.79
2024-Q1	0.6552	0.7917	0.6769	0.9533	189.02
2024-Q2	0.7804	0.8767	0.9655	0.8028	−96.03
2024-Q3	0.5669	0.7236	0.5703	0.9894	395.49
2024-Q4	0.6496	0.7876	0.7263	0.8601	124.49
2025-Q1	0.8740	0.9328	0.9362	0.9293	−4.55
2025-Q2	0.9407	0.9695	0.9583	0.9809	14.04
2025-Q3	0.5745	0.7297	0.8544	0.6368	−254.68
2025-Q4	0.7061	0.8278	0.8583	0.7994	−64.14
Mean	0.6972	0.8170	0.8054	0.8511	32.47
Std	0.1043	0.0725	0.1111	0.1004	188.53

TABLE X. PERFORMANCE COMPARISON OF ALL ARCHITECTURES ON BI-MONTHLY DATA.

Model	Setup	IoU	Dice	P	R	$\lvert\Delta A\rvert$
ConvLSTM	S1	0.6347	0.7732	0.7941	0.7708	131.44
ConvLSTM	S2	0.6623	0.7956	0.7979	0.8062	119.11
U-Net+LSTM	S1	0.7439	0.8435	0.8658	0.8496	156.64
U-Net+LSTM	S2	**0.7791**	**0.8701**	**0.8971**	0.8691	160.78
Att.-U-Net+CLSTM	S1	0.7410	0.8406	0.8806	0.8293	157.45
Att.-U-Net+CLSTM	S2	0.7733	0.8662	0.9094	0.8459	**147.45**
Swin	S1	0.6333	0.7718	0.8015	0.7608	131.99
Swin	S2	0.6952	0.8180	0.8404	0.8086	114.45
ViT	S1	0.3805	0.5487	0.5461	0.5771	203.91
ViT	S2	0.3421	0.5079	0.4723	0.5711	243.07

Attention U-Net+ConvLSTM dominance at yearly/quarterly resolutions to U-Net+LSTM leadership at bi-monthly suggests spatial feature extraction quality becomes progressively more critical as temporal frequency increases (see Table X).

G. Long-Term Morphological Forecasting (2026-2040)

In long-term morphological forecasting, final model setup was Attention U-Net+ConvLSTM configuration trained on complete dataset covering time span 1987-2025 with a window length equal to five. Convergence point for training achieved epoch 107, best validation loss being equal to 0.2298, validation Dice index 0.8057, and validation IoU 0.7026. Reconstruction test for known time interval (2016–2025) revealed mean IoU = 0.6543 and mean Dice index 0.7907, thus indicating that model is capable of reconstructing historical river extent patterns sufficiently accurately.

Starting from the 2025 area of 517.6 km^2, river extent grew up gradually to 671.3 km^2 in 2026 and reached value of 807.6 km^2 by 2040, indicating total growth of 290.0 km^2 or 56.03% as compared to the base year. Expansion trend had nonuniform character, showing largest growth rates during the first years of forecasting (2025-2026) and further decreasing gradually. The accretion-dominated pattern indicates future dynamics primarily associated with lateral expansion into adjacent floodplain areas rather than wholesale channel widening.

The comprehensive long-term risk cartography (Fig. 6) identifies spatially explicit high-probability morphological change zones concentrated systematically along active channel margins and proximal floodplain corridor segments, with very-high-risk designations indicating persistent instability and recurrent channel boundary adjustment patterns most operationally relevant for infrastructure protection planning and settlement management strategies.

H. Short-Term Bi-Monthly Forecasting (2026-2028)

Short-term high-frequency bi-monthly forecasting employed U-Net+ConvLSTM2D architecture with nine-period input sequences systematically enriched with sinusoidal and cosinusoidal seasonal encoding channels ($\sin(2\pi t/6)$, $\cos(2\pi t/6)$ where t represents bi-monthly period index), enabling model to explicitly distinguish intra-annual temporal periods without relying exclusively on spatial morphological content patterns. Training convergence occurred at epoch 39 (validation Dice = 0.8381, IoU = 0.7463). Reconstruction sanity check on recent historical period 2023 P1–2025 P6 yielded mean IoU = 0.684 ± 0.072 (mean $\pm$ std).

Over 18 forecast periods spanning 2026 P1–2028 P6, mean predicted area followed clear three-phase temporal structure: initial contraction phase to mean 902.7 km^2 in 2026 (below 934.1 km^2 baseline), reflecting dry-to-wet seasonal transition inherited from seed sequence; progressive rise to 984.0 km^2 in 2027; further increase to 1,020.1 km^2 in 2028, with terminal forecast period (2028 P6, November–December) reaching 1,054.4 km^2 (+12.88% above baseline). The seasonal cycle was faithfully reproduced with characteristic pattern: minimum extent at P3 (May–June pre-monsoon, 918.3 km^2), maximum at P5 (September–October peak monsoon, 1,034.3 km^2), seasonal amplitude 116.0 km^2, and mean monsoon area 53.4 km^2 (+5.62%) above non-monsoon mean. Final forecast period exhibited accretion-to-erosion ratio approximately 1.60:1, with 78.4% of baseline channel extent maintaining stable classification.

I. Computational Complexity

Table XI summarizes the model complexity in terms of trainable parameters and average training time across different temporal resolutions.

J. Forecast Reliability and Uncertainty

Long-term forecasting is subject to uncertainty because prediction errors can accumulate over successive time steps. To assess forecast reliability, rolling predictions were evaluated across two controlled experimental setups: a 10-year horizon (Setup 1: 2016–2025) and a 5-year horizon (Setup 2: 2021–2025). The best-performing Attention U-Net+ConvLSTM model maintained remarkably stable performance across these testing intervals, achieving a mean IoU of 0.7005 (Dice = 0.8236) in the 10-year long-range forecast setup, and a mean IoU of 0.6963 (Dice = 0.8205) in the 5-year medium-range setup.

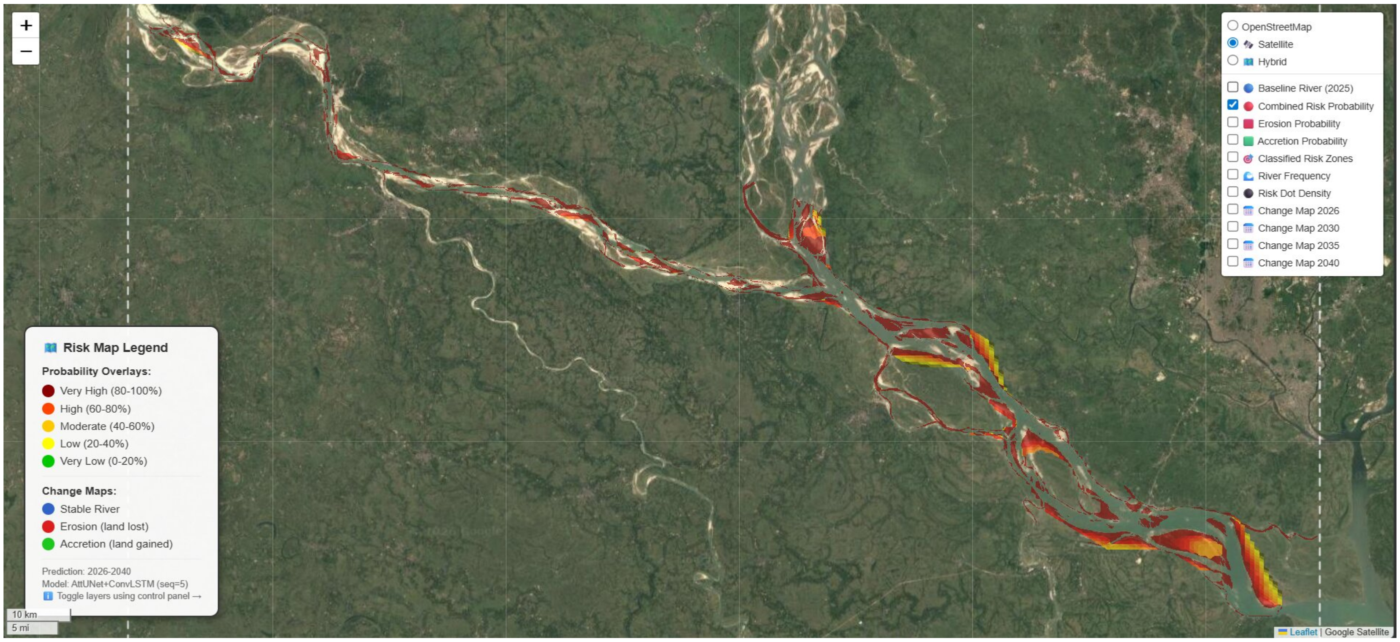

Fig. 6. Long-term combined morphological risk for 2026-2040. Interactive web demonstration available here.

TABLE XI. Model Complexity in Terms of Trainable Parameters, Training Time, and Inference Time Per Prediction.

Model	Data	Params (M)	Train (min)	Infer (s)
ConvLSTM	Yearly	1.41	8.2	0.31
	Quarterly		34.5	0.50
	Bi-monthly		54.1	0.65
U-Net + LSTM	Yearly	7.71	12.4	0.71
	Quarterly		48.2	1.10
	Bi-monthly		72.5	1.30
Attention U-Net + ConvLSTM	Yearly	**7.41**	**14.1**	**0.81**
	Quarterly		**52.8**	**1.10**
	Bi-monthly		**79.3**	**1.30**
ViT	Yearly	3.24	6.5	0.40
	Quarterly		22.0	0.62
	Bi-monthly		48.2	0.83
Swin Transformer	Yearly	5.07	22.0	0.60
	Quarterly		78.3	0.80
	Bi-monthly		115.4	1.00

Furthermore, the model exhibited tighter dimensional calibration over the shorter testing window, yielding a lower mean absolute area difference ($|\Delta A| = 41.05$ km^2) in Setup 2 compared to Setup 1 ($|\Delta A| = 66.54$ km^2). This behavior confirms that shorter horizons effectively curb cumulative error propagation. Rather than presenting future channel changes via strict binary boundaries, long-term channel evolution (2026–2040) is mapped using spatially explicit risk zones derived from pixel-level water occurrence frequencies (F_w) and a hydrodynamic stability index (I). This approach accounts for spatial modeling uncertainties and highlights highly dynamic zones along active channel margins.

VI. Discussion

Hybrid CNN-LSTM architectures outperformed transformer-based architectures in terms of overlap and area metrics for all datasets and forecasting horizons. Underperformance of transformer-based architectures (ViT, Swin Transformer) is presumably related to data scarcity. The limited temporal samples (38 yearly, 125 quarterly, and 156 bi-monthly composites) likely increased the risk of overfitting for attention-based models. Although a full hyperparameter sweep was not conducted, key architecture-specific parameters (e.g., patch size, window size, and embedding dimension) were determined via iterative manual testing, and the most effective configuration was selected for final evaluation. Hierarchical window-based attention introduces stronger spatial locality biases, which may explain its better performance than ViT. With growing satellite databases and availability of pre-trained geospatial foundation models, transformer architectures are anticipated to be competitive with CNN-LSTM hybrids.

The Attention U-Net+ConvLSTM was best suited for yearly and quarterly forecasting, where attention helps focus on geographically active regions. The U-Net+LSTM architecture was superior for bi-monthly forecasts when boundary preservation is crucial but where gating provided only incremental benefits.

There was a significant seasonal influence on the performance of networks in quarter- and bi-monthly experiments. It relates to the difficulty associated with predicting the monsoon extent, where channel widths become large, have unusual forms, and become inundated with floodplain waters. For Setup 1 quarterly predictions, mean areas were close to each other (–0.47 km²), suggesting annual canceling of seasonal systematic errors.

An essential realization in regard to operational use is that metrics like Intersection Over Union (IoU) and Dice score are insufficient measures of model performance. Several model configurations had similar metrics of overlap but vastly differing absolute area differences, showing spatially accurate but incorrectly dimensionally sized predictions. River hazard management processes such as erosion risk calculations, land lost calculations in agriculture, embankments construction, and disaster resource allocations require accurate area calculations. It is recommended that IoU, Dice, Precision, Recall, and absolute area difference be considered minimum requirements for performance metrics, with emphasis on area difference in selection process for hazard management.

Future prediction from 2026 to 2040 shows continuing expansion with accretion outpacing erosion. This corresponds with past observations between 2008 and 2018 (net accretion of 113.0 km²) and near equilibrium 2018 to 2025, indicating models learned realistic trends for the future. The slowing expansion beyond 2031 may indicate learned stabilizing effects; however, potential cumulative errors in modeling may be responsible. Historical reconstruction having consistent positive area bias (91.8 km²) indicates that the expansion values calculated are upper limit estimates. Nevertheless, the more accurate risk zone maps defining corridor segments with high change probability provide the main output for disaster management.

A. Limitations and Future Research Directions

However, several limitations need mentioning. Firstly, 60 m Landsat spatial resolution fails to capture erosion micro-features, secondary channels, and banks' morphology, thus requiring shifting to 10 m Sentinel-2 spatial resolution data through the whole period of analysis. Secondly, lack of hydrological forcing data in the form of river discharge, precipitation, and antecedent soil moisture presents a serious limitation for the framework under development, especially during the periods of flooding. Thirdly, the proposed approach lacks any geotechnical parameters such as the soil type, bank height, and vegetation distribution, which influence erosion processes. Lastly, all the presented experiments were done on only one river system, making cross-system testing on Jamuna and Meghna river systems necessary to make any confident generalizations.

The following directions may be considered in future research: (1) inclusion of 10 m Sentinel-2 optical and Sentinel-1 SAR data through the entire time series; (2) widening the geographic scope to other river systems – Jamuna and Meghna; (3) including hydrological variables as forcing data to better predict the extremes; (4) exploring transformer neural networks pre-trained on larger geospatial foundation models corpora; and (5) combining geomorphology-based vulnerability map with population density and agriculture maps for a socio-economic early warning system.

VII. Conclusion

This investigation presents a comprehensive framework for analyzing and forecasting morphological evolution of the Padma River by combining long-term satellite remote sensing with spatiotemporal deep learning. A 38-year satellite-derived dataset enabled comprehensive analysis of long-term river dynamics, revealing substantial erosion, accretion, and channel migration. The proposed preprocessing framework produced reliable historical water masks for subsequent forecasting. Hybrid CNN-LSTM architectures consistently outperformed transformer-based models under limited training data. The proposed framework accurately captured both long-term morphological evolution and short-term seasonal dynamics, enabling spatially explicit future risk assessment. This research demonstrates that freely available satellite imagery combined with cloud-based processing and spatiotemporal deep learning can provide a practical, scalable, and reproducible framework for riverbank hazard monitoring and multi-year forecasting. Furthermore, this remote sensing and deep learning pipeline is fundamentally generic and can be seamlessly applied to any other dynamic river system by retraining the models on their respective historical data. For those living near erodible banks along rivers throughout Bangladesh and South Asia, predictive tools capable of identifying hazardous areas can meaningfully enhance disaster preparedness, infrastructure planning, and community adaptation strategies.

References

[1] P. Lemenkova, "Deep learning methods of satellite image processing for monitoring of flood dynamics in the Ganges delta, Bangladesh," *Water*, vol. 16, no. 8, p. 1141, 2024.

[2] J. Freihardt *et al.*, "Assessing riverbank erosion in Bangladesh using time series of Sentinel-1 radar imagery in the Google Earth Engine," *Natural Hazards and Earth System Sciences*, vol. 23, pp. 751–768, 2023.

[3] G. M. M. Alam *et al.*, "Exploring impacts and livelihood vulnerability of riverbank erosion hazard among rural household along the river Padma of Bangladesh," *Environmental Systems Research*, vol. 6, p. 25, 2017.

[4] M. N. Islam *et al.*, "Socioeconomic impacts and migration dynamics of riverbank erosion," *npj Climate Action*, vol. 4, p. 6, 2025.

[5] T. K. Das *et al.*, "Impact of riverbank erosion: A case study," *Australasian Journal of Disaster and Trauma Studies*, vol. 21, no. 2, pp. 73–81, 2017.

[6] D. Sarkar *et al.*, "Assessing riverbank erosion and livelihood resilience using traditional approaches in northern Bangladesh," *Sustainability*, vol. 14, no. 4, p. 2348, 2022.

[7] H. Fan *et al.*, "Using long short-term memory networks for river flow prediction," *Hydrology Research*, vol. 51, no. 6, pp. 1358–1376, 2020.

[8] A. Dehghani *et al.*, "Comparative evaluation of LSTM, CNN, and ConvLSTM for spatiotemporal forecasting tasks," *Ecological Informatics*, vol. 75, p. 101984, 2023.

[9] D. V. Binh *et al.*, "A novel method for river bank detection from Landsat time series (SRBED)," *Remote Sensing*, vol. 12, no. 20, p. 3298, 2020.

[10] T. Langhorst *et al.*, "Global observations of riverbank erosion and accretion from Landsat imagery," *Journal of Geophysical Research: Earth Surface*, vol. 128, no. 2, p. e2022JF006774, 2023.

[11] P. K. Langat *et al.*, "Monitoring river channel dynamics using remote sensing and GIS techniques," *Geomorphology*, vol. 325, pp. 92–106, 2019.

[12] M. N. Islam and M. A. Haque, "River channel migration: A remote sensing and GIS analysis," in *Proc. ESA Living Planet Symposium*, Bergen, Norway, 2010.

[13] P. V. Raju *et al.*, "Spatio-temporal analysis of riverbank changes using remote sensing and geographic information system," *Ecological Informatics*, vol. 81, p. 102628, 2024.

[14] T. Tha, T. Piman, D. Bhatpuria, and P. Ruangrassamee, "Assessment of riverbank erosion hotspots along the Mekong River in Cambodia using remote sensing and hazard exposure mapping," *Water*, vol. 14, no. 19, p. 2995, 2022.

[15] O. R. Saha, "Riverbank migration and island dynamics at the Padma confluence: Multi-sensor analysis," *Geomorphology*, vol. 448, p. 109043, 2025.

[16] M. A. Haque, "Change detection of Jamuna River and its impact on local land use and settlement," *Journal of Urban and Regional Analysis*, vol. 15, no. 2, pp. 145–162, 2023.

[17] S. M. Ritu, S. K. Sarkar, and H. Zonaed, "Prediction of Padma river bank shifting and its consequences on LULC changes," *Ecological Indicators*, vol. 156, p. 111104, 2023.

[18] C. Muzahid *et al.*, "Quantifying river bank erosion and accretion patterns along the Gorai River in Kushtia, Bangladesh," *Journal of Geographic Information System*, vol. 16, pp. 70–88, 2024.

[19] Z. Ren, "A large-scale riverbank erosion risk assessment model with remote sensing inputs," *Applied Geography*, vol. 162, p. 103156, 2024.

[20] K. Yuan *et al.*, "Deep-learning-based multispectral satellite image segmentation for water body detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 7422–7434, 2021.

[21] Y. Li *et al.*, "Assessing the performance and interpretability of the CNN-LSTM-Attention model for daily streamflow forecasting in typical basins of the eastern Qinghai-Tibet Plateau," *Scientific Reports*, vol. 15, p. 260, 2025.

[22] S. Ghosh *et al.*, "Water body extraction from high spatial resolution remote sensing images based on enhanced U-Net and multi-scale information fusion," *Scientific Reports*, vol. 14, p. 16192, 2024.

[23] B. DeVries *et al.*, "Rapid and robust monitoring of flood events using Sentinel-1 and Landsat data on the Google Earth Engine," *Remote Sensing of Environment*, vol. 240, p. 111664, 2020.

[24] M. A. Clement *et al.*, "Visualisation of flooding along an unvegetated, ephemeral river using Google Earth Engine," *Journal of Environmental Management*, vol. 277, p. 111338, 2020.

[25] P. Tripathy and A. Mallick, "Applying Google Earth Engine for flood mapping and monitoring in the downstream provinces of Mekong River," *Results in Engineering*, vol. 14, p. 100432, 2022.

[26] N. Sazib *et al.*, "Google Earth Engine for large-scale flood mapping using SAR data and impact assessment on agriculture and population of Ganga-Brahmaputra basin," *Sustainability*, vol. 14, no. 7, p. 4210, 2022.

[27] X.-H. Le *et al.*, "Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks," *Scientific Reports*, vol. 11, p. 17429, 2021.

[28] Z. Xiang, J. Yan, and I. Demir, "Hybrid CNN-LSTM models for river flow prediction," *Water Supply*, vol. 22, no. 5, pp. 4902–4919, 2022.

[29] A. Horvath *et al.*, "Water level prediction using long short-term memory neural network model for a lowland river: A case study on the Tisza River, Central Europe," *Environmental Sciences Europe*, vol. 35, p. 92, 2023.

[30] Z. Liu *et al.*, "Video Swin Transformer," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 3202–3211.

[31] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. ICLR*, 2021.

[32] A. Arnab *et al.*, "ViViT: A video vision transformer," in *Proc. IEEE/CVF ICCV*, 2021, pp. 6836–6846.

[33] J. Garber *et al.*, "Artificial intelligence and objective-function methods can identify bankfull river channel extents," *Water Resources Research*, vol. 60, no. 1, p. e2023WR035269, 2024.

[34] H. B. Mann, "Nonparametric tests against trend," *Econometrica*, vol. 13, no. 3, pp. 245–259, 1945.

[35] M. G. Kendall, *Rank Correlation Methods*, 4th ed. London: Charles Griffin, 1975.

[36] H. Theil, "A rank-invariant method of linear and polynomial regression analysis," *Proceedings of the Royal Netherlands Academy of Sciences*, vol. 53, pp. 386–392, 1950.

[37] P. K. Sen, "Estimates of the regression coefficient based on Kendall's tau," *Journal of the American Statistical Association*, vol. 63, no. 324, pp. 1379–1389, 1968.

[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2015.