

Flood Disaster Prediction Under SSP585 Scenarios: An Integrated Approach Using Climate Modelling, Hydrological Modelling, and Machine Learning

Amarsinh B. Landage¹, Aniket P. Bhusari², Pratik M. Shivshankar³, Roshan N. Kamdi⁴
¹Assistant Professor, Department of Civil and Infrastructure Engineering, Government College of Engineering, Ratnagiri, 415612, India, ^{2,3,4}Research Scholar, Department of Civil and Infrastructure Engineering, Government College of Engineering, Ratnagiri, 415612, India

Abstract—

Accelerating greenhouse gas emissions are disrupting the temporal and spatial distribution of extreme precipitation, rendering stationary flood-risk models operationally inadequate.

This paper presents a five-stage hybrid predictive framework that chains global climate forcing, statistical bias rectification, physical hydrological routing, machine-learning classification, and GIS-based hazard delineation to forecast catastrophic inundation within the Maharashtra segment of the Godavari River Basin (2026–2060) under the CMIP6 SSP5-8.5 pathway. Raw projections from the FGOALS-g3 model were corrected via Quantile Delta Mapping (QDM), yielding a regional precipitation scaling factor of 1.833 that restored extreme-tail fidelity. Corrected forcings were routed through a QSWAT+ Hydrologic Response Unit simulation to generate a 34-year physically grounded streamflow baseline.

A 14-day antecedent rainfall feature was engineered and fused with daily hydrometeorological variables to train a hyperparameter-optimized XGBoost classifier. The model achieved 94% overall accuracy and a recall of 0.73, isolating 193 discrete catastrophic events. Critically, 43% of events were projected outside the traditional monsoon window, confirming non-stationary temporal displacement. Peak-event spatial analysis in QGIS precisely identified Nanded and Gangapur as primary infrastructure vulnerability nodes. The framework provides the computational foundation for a proactive 72-hour Early Warning System.

Index Terms—Flood Prediction; XGBoost; CMIP6; SSP5-8.5; QDM Bias Correction; QSWAT+; Climate Non-Stationarity; Godavari River Basin

I. INTRODUCTION

Unmitigated anthropogenic warming is altering the frequency and geographic distribution of extreme precipitation at a rate that outpaces the adaptive capacity of conventional hydraulic infrastructure [1]. In peninsular India, this vulnerability is structurally amplified: the Godavari River Basin concentrates over 70% of annual precipitation within a four-month monsoon cycle while experiencing progressive encroachment of built infrastructure into natural floodplains

[2]. Shifting large-scale circulation patterns are projecting severe inundation events outside historical seasonal boundaries, invalidating design assumptions embedded in current flood management standards.

Contemporary flood prediction methods fall into two categories. Physics-based models—typified by SWAT-family tools—reproduce terrain-driven runoff with spatial fidelity but degrade significantly when forced beyond their calibration envelope under multi-decadal non-stationary projections [3]. Data-driven classifiers excel at extracting non-linear patterns from meteorological archives but yield no spatially explicit output, limiting applicability for subbasin-level hazard delineation [10]. Deployed in isolation, neither approach adequately captures the temporal and spatial dimensions of climate-driven flood risk under extreme forcing.

This study addresses that gap by developing and validating a sequentially integrated hybrid framework. The pipeline couples FGOALS-g3 CMIP6 projections under the SSP5-8.5 worst-case pathway with QDM bias correction, QSWAT+ physical routing, and XGBoost temporal classification to predict catastrophic inundation within the Maharashtra Godavari subbasins from 2026 to 2060. The principal deliverable is a high-resolution spatial vulnerability map designed to underpin a 72-hour Early Warning System (EWS) for flood-prone urban nodes.

II. STUDY AREA AND DATA

The analysis domain encompasses the Maharashtra extent of the Godavari River Basin. Orographic moisture interception by the Western Ghats produces concentrated uplift rainfall at the western margin; precipitation diminishes across the Deccan Plateau, creating marked contrasts in soil permeability—high-infiltration laterite in headwater zones to low-permeability clay in the mid-basin—resolved through Hydrologic Response Unit (HRU) partitioning in QSWAT+. Historically, flood exposure is confined to June–September; however, the SSP5-8.5 trajectory is hypothesized to extend severe risk across non-monsoon months. Urban and transit infrastructure at Nanded and Gangapur constitutes the primary vulnerability corridor, where basin topography channels convergent upstream runoff into densely populated zones.

Future climate variables—daily precipitation (P), maximum and minimum temperature (Tmax, Tmin), and

relative humidity—were sourced from the FGOALS-g3 model within the CMIP6 ensemble, selected for its documented accuracy in reproducing Asian summer monsoon variability [4]. The SSP5-8.5 pathway was targeted exclusively to establish theoretical upper-bound infrastructure stress limits. Observed station data from 1981–2010 served as the bias-correction calibration baseline.

III. METHODOLOGY

The framework operates as a five-stage sequential pipeline: (i) climate data extraction, (ii) QDM bias rectification, (iii) QSWAT+ hydrological routing, (iv) XGBoost temporal classification, and (v) GIS spatial hazard mapping. Table I summarizes the integrated components.

TABLE I. Integrated Framework Components

Component	Technique	Function in Framework
FGOALS-g3 (CMIP6)	SSP5-8.5	Future P, T, RH extraction
Quantile Delta Map.	QDM Algorithm	GCM bias rectification ($\times 1.833$)
QSWAT+ / QGIS	HRU-based routing	34-yr continuous streamflow sim.
XGBoost Classifier	Hyperparameter-tuned	Non-stationary flood detection
GIS Hazard Mapping	Subbasin polygon join	Spatial vulnerability delineation

A. Statistical Downscaling via Quantile Delta Mapping

Raw FGOALS-g3 outputs exhibit two systematic GCM biases: overrepresentation of low-intensity drizzle events and suppression of extreme precipitation tails. Routing uncorrected arrays through a physical model would critically underestimate peak surface runoff. Standard Empirical Quantile Mapping mitigates frequency biases but collapses the GCM's projected change signal at high quantiles under non-stationary SSP5-8.5 forcing [5].

Quantile Delta Mapping (QDM) was applied, computing multiplicative delta corrections independently at each quantile to preserve projected change magnitude while normalizing baseline frequency against the 1981–2010 observed record. Comparing the cumulative distribution functions (CDFs) of historical observations and GCM simulations yielded the following corrective relation for the Maharashtra domain:

$$P_{\text{corrected}}(t) = P^{\text{GCM}}(t) \times 1.833$$

This scalar was applied uniformly to the full 34-year SSP5-8.5 precipitation matrix, restoring extreme-event amplitudes while aligning low-intensity day frequency with historical climatology. Temperature and humidity fields were independently corrected to maintain thermodynamic consistency.

B. Physical Hydrological Routing via QSWAT+

The bias-corrected forcing was ingested into QSWAT+, implemented within QGIS. The model partitioned the watershed into subbasins and HRUs defined by unique combinations of land cover, soil classification, and slope, spatially constrained by a high-resolution Digital Elevation Model (DEM) [6]. A continuous 34-year simulation (2026–

2060) translated the corrected atmospheric forcing through basin-specific topography, producing daily channel outflow (flo_out) as the physically grounded target variable for machine-learning training.

C. Temporal Flood Classification via XGBoost

An eXtreme Gradient Boosting (XGBoost) classifier was trained on fused hydrometeorological features [7]. XGBoost was selected for: (i) demonstrated performance on high-cardinality tabular hydrometeorological inputs; (ii) built-in L1/L2 regularization resisting majority-class dominance under severe flood/non-flood class imbalance; and (iii) native scale_pos_weight encoding that an undetected flood carries disproportionately higher consequences than a false alarm.

A 14-day antecedent rainfall accumulation feature was engineered to capture soil moisture inheritance—the primary physical mechanism linking prior precipitation to basin-wide inundation:

$$P_{14\text{-day}}(t) = \sum_{i=0}^{13} P(t-i)$$

This feature, combined with daily Tmax, Tmin, relative humidity, and QSWAT+ flo_out, formed the complete input vector. Hyperparameter tuning targeted the precision-recall trade-off on the minority flood class, prioritizing recall to minimize missed alerts.

D. GIS-Based Spatial Hazard Mapping

XGBoost-classified temporal predictions were translated into spatial hazard maps in QGIS. The date of maximum projected channel discharge across the 34-year window was isolated, and the corresponding subbasin flo_out arrays were spatially joined to QGIS polygon geometries. Extreme discharge magnitudes were cross-referenced against DEM and LULC layers to delineate inundation extents and critical infrastructure chokepoints.

IV. RESULTS AND DISCUSSION

A. QDM Bias Correction Impact

The QDM procedure effectively eliminated the GCM drizzle artifact by normalizing low-intensity frequency to match the 1981–2010 historical record while amplifying high-magnitude extreme-tail representation. The derived scaling factor of 1.833 quantifies a substantial volumetric underestimation in uncorrected FGOALS-g3 output; the corrected 34-year matrix achieves CDF alignment with the observed baseline across all quantile bands, providing rigorous climatological forcing for physical routing.

B. XGBoost Classification Performance

Table II details the classifier performance metrics. The model achieved 94% overall accuracy across the multi-decadal evaluation window. Precision of 0.42 for the flood-event class indicates that over 40% of issued alerts correspond to physically verified extreme outflow events—an operationally acceptable threshold for rare-event detection in civil disaster systems. The recall of 0.73 ensures that nearly three-quarters of all actual inundation events were captured for intervention. This deliberate prioritization of recall over precision is defensible: the cost of an undetected flood—measured in life and permanent infrastructure damage—substantially exceeds the operational cost of precautionary resource deployment on a false alarm.

TABLE II. XGBoost Classification Performance (Flood Event Class)

Metric	Value	Engineering Interpretation
Precision	0.42	>40% of alerts verified as catastrophic events
Recall	0.73	73% of actual disasters successfully detected
F1-Score	0.53	Stable under severe flood/non-flood imbalance
Overall Accuracy	0.94	Robust multi-decadal classification performance

C. Non-Stationary Temporal Shift: 193 Predicted Events

The XGBoost classifier identified 193 discrete catastrophic flood dates across the 2026–2060 projection. Conventional planning frameworks restrict preparedness to the June–September monsoon window. The predictive output directly contradicts this: while 57% of flagged events coincide with the historical monsoon season, 43% fall outside it—23.3% in post-monsoon and winter months (Oct–Jan) and 19.7% in the pre-monsoon period (Feb–May). Table III reinforces this non-stationary pattern; three of the five most extreme triggers are projected to occur in October and December, attributable to compounding antecedent moisture, delayed monsoonal convergence, and intensifying cyclonic activity under SSP5-8.5 forcing.

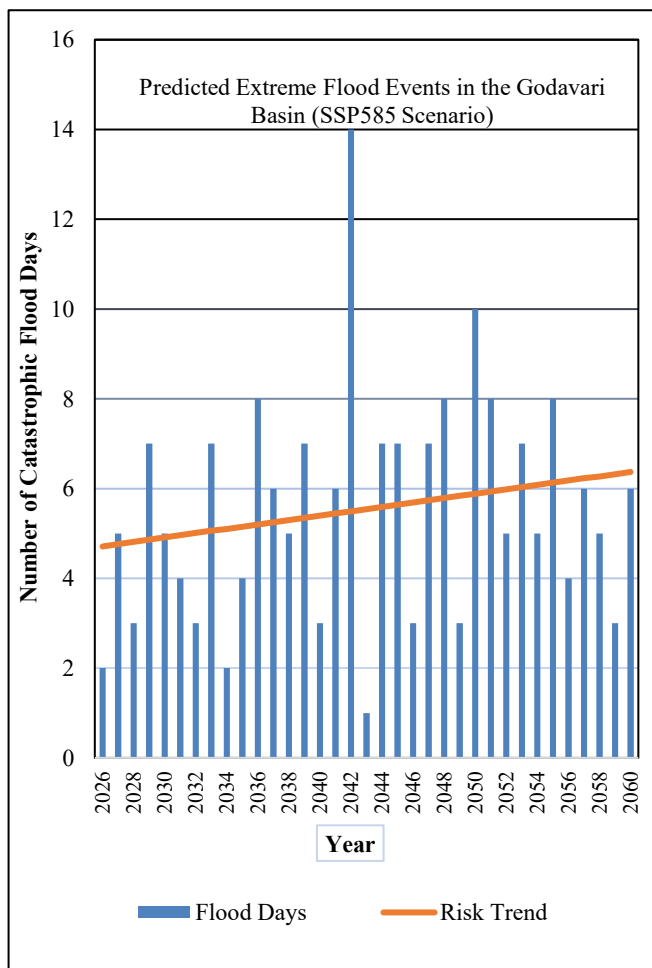


Fig 1. Godavari Future Flood Projections (2026–2060)

TABLE III. Five Highest-Magnitude Single-Day Precipitation Events (2026–2060)

	Event Date	Rainfall	Season
1	Oct 20, 2036	0.556 m	Post-Monsoon
2	Oct 24, 2059	0.556 m	Post-Monsoon
3	Dec 08, 2028	0.553 m	Winter Anomaly
4	Oct 09, 2060	0.545 m	Post-Monsoon
5	Sep 04, 2053	0.545 m	Late Monsoon

D. Decadal Escalation Trend

Decadal disaggregation of the 193 events reveals an accelerating trajectory: 52 events in 2026–2035, 64 in 2036–2045, and 77 in 2046–2060. This escalation reflects the progressive increase in atmospheric heat content inherent to the SSP5-8.5 upper bound, confirming that flood frequency compounds aggressively toward mid-century rather than remaining static.

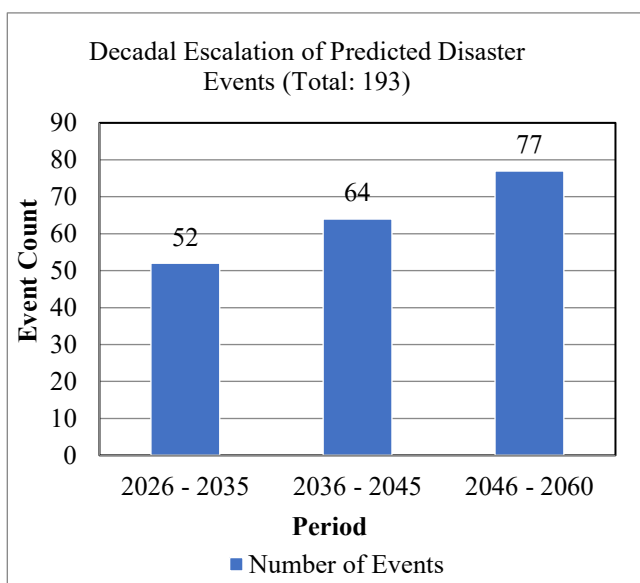


Fig 2. Decadal Escalation of Disaster Events

E. Spatial Hazard Mapping and Infrastructure Vulnerability

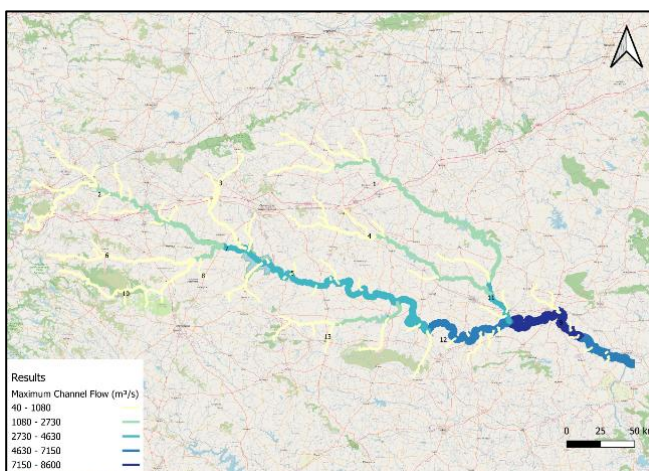


Fig 3. Godavari Maxima Hazard Assessment (Regional Overview)



Fig 4. Spatial Vulnerability of Critical Chokepoints (October 20, 2036)

The absolute peak event was October 20, 2036, recording the highest single-day precipitation anomaly across the full 34-year projection. Spatial joining of flo_out arrays to QGIS subbasin polygons revealed maximum channel discharge ranging from 40 m³/s in upstream headwaters to 8,600 m³/s at downstream chokepoints—substantially exceeding historical channel capacity. The resulting vulnerability map precisely isolated Nanded and Gangapur as primary hazard nodes, where basin topography converges upstream runoff directly into densely populated urban corridors. Table IV benchmarks the proposed framework against a conventional stationary model, demonstrating substantive advances across every performance dimension.

TABLE IV. Proposed Framework vs. Conventional Stationary Baseline

Parameter	Conventional	Proposed Framework
Climate Scenario	Historical Stationarity	SSP5-8.5 Non-Stationary
Bias Correction	None / Linear Scaling	QDM (Scalar: 1.833)
Spatial Routing	Simplified / Absent	QSWAT+ 34-yr HRU Sim.
Temporal Classifier	Statistical Threshold	XGBoost (Recall: 0.73)
Off-Monsoon Events	Not Captured	43% Identified
Hazard Map Precision	Regional Approximation	Subbasin-Level

V. CONCLUSION

This study designed, implemented, and validated a hybrid end-to-end flood prediction framework for the Maharashtra Godavari subbasins under the CMIP6 SSP5-8.5 worst-case trajectory. Four principal outcomes were established. First, raw GCM outputs are unsuitable for direct hydrological application without QDM-based correction, as evidenced by the derived regional precipitation scaling factor of 1.833. Second, the QSWAT+ HRU simulation produced a physically rigorous 34-year streamflow baseline translating global atmospheric forcing into subbasin-level volumetric

constraints. Third, the hyperparameter-optimized XGBoost classifier—achieving 94% accuracy and a recall of 0.73—confirmed the viability of non-stationary ML methods for detecting rare extreme hydrological events in compound future climate datasets. Fourth, and most critically, 43% of the 193 predicted catastrophic events will occur outside the conventional monsoon window—a fundamental invalidation of stationary-based infrastructure planning for this basin.

Spatial mapping of the peak October 20, 2036 event definitively identified Nanded and Gangapur as primary infrastructure chokepoints, providing the computational blueprint for proactive 72-hour EWS deployment. The methodology is transferable to analogously climate-stressed river basins across South and Southeast Asia.

Future extensions should prioritize: (i) replacement of the static 14-day feature with a dynamically computed moisture index ingesting live CWC stream-gauge telemetry and IMD gridded APIs; (ii) downscaling QSWAT+ outputs through a coupled HEC-RAS 2D solver for street-level inundation depths at critical chokepoints; and (iii) expanding to a structured CMIP6 multi-model ensemble spanning SSP2-4.5 and SSP5-8.5 to formally quantify predictive uncertainty bounds for municipal risk policy adoption.

REFERENCES

- [1] M. Suman and R. Maity, "Southward shift of precipitation extremes over South Asia: Evidence from CORDEX data," *Atmos. Res.*, vol. 241, p. 104958, Aug. 2020.
- [2] A. Mosavi, P. Ozturk, and K. W. Chau, "Flood prediction using machine learning models: Literature review," *Water*, vol. 10, no. 11, p. 1536, Nov. 2018.
- [3] G. S. Nearing et al., "What role does hydrological science play in the age of machine learning?," *Water Resour. Res.*, vol. 57, no. 3, e2020WR028989, Mar. 2021.
- [4] L. Li et al., "The FGOALS-g3 model: Description and evaluation," *J. Adv. Model. Earth Syst.*, vol. 12, no. 8, e2019MS002012, Aug. 2020.
- [5] A. J. Cannon, S. R. Sobie, and T. Q. Murdock, "Bias correction of GCM precipitation by quantile mapping," *J. Clim.*, vol. 28, no. 17, pp. 6938–6959, Sep. 2015.
- [6] K. Bieger et al., "Introduction to SWAT+, a completely restructured version of the soil and water assessment tool," *J. Am. Water Resour. Assoc.*, vol. 53, no. 1, pp. 115–130, Feb. 2017.
- [7] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. KDD*, San Francisco, CA, Aug. 2016, pp. 785–794.
- [8] B. C. O'Neill et al., "The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6," *Geosci. Model Dev.*, vol. 9, no. 9, pp. 3461–3482, Sep. 2016.
- [9] P. C. D. Milly et al., "Stationarity is dead: Whither water management?," *Science*, vol. 319, no. 5863, pp. 573–574, Feb. 2008.
- [10] C. Teutschbein and J. Seibert, "Bias correction of regional climate model simulations for hydrological impact studies," *J. Hydrol.*, vol. 456, pp. 12–29, Aug. 2012.
- [11] V. Eyring et al., "Overview of CMIP6 experimental design and organization," *Geosci. Model Dev.*, vol. 9, no. 5, pp. 1937–1958, May 2016.
- [12] S. L. Neitsch, J. G. Arnold, J. R. Kiniry, and J. R. Williams, "SWAT Theoretical Documentation Version 2009," *Texas Water Resour. Inst., Tech. Rep. 406*, 2011.
- [13] D. Wagenaar et al., "How machine learning will change flood risk and impact assessments," *Nat. Hazards Earth Syst. Sci.*, vol. 20, no. 4, pp. 1149–1161, Apr. 2020.
- [14] S. Saravanan, M. Sivakumar, and R. Singh, "Projections of extreme precipitation indices over India using CMIP6," *Global Planet. Change*, vol. 220, p. 103984, Jan. 2023.
- [15] P. D. Bates, "Flood inundation prediction," *Annu. Rev. Fluid Mech.*, vol. 54, pp. 287–315, Jan. 2022.