

AUTHOR RESPONSE TO REFEREE COMMENTS

Manuscript egusphere-2026-1275

Hydrologically Enhanced Machine Learning Framework for Urban Flood
Inundation Mapping Using Multi-Sensor Remote Sensing Data: A Case Study of
Mumbai, India

Authors: Ankush S. Pawar and Gayatri M. Phade

Response to the Editor and Referees

We sincerely thank the Executive Editor and all five referees for their careful assessment and constructive comments. We have revised the manuscript substantially in response to the reports. The principal revisions concern clarification of the original-research contribution, the formulation and interpretation of REM and RNI, threshold-based consensus flood-inventory generation, target–predictor dependence, validation limitations, statistical assessment, comparison with previous studies, and moderation of claims concerning generalisation and operational use.

Important note: Where an independent observation set was not available, we have not presented the available historical-location comparison as independent quantitative validation. The revised manuscript explicitly states this limitation.

EDITORIAL / MANUSCRIPT-TYPE CLARIFICATION

Comment: The manuscript was classified as a “Review article”, although it presents an original methodological framework, data processing, flood-inventory generation, machine-learning experiments and results.

Response: We agree. The work is an original research study. The revised manuscript is structured and written as an original research contribution. We respectfully request correction of the manuscript type in the editorial/production metadata to the appropriate original-research category, if this can be changed during production.

Change made: The title page and manuscript wording no longer present the work as a review. The final production files will follow the required author–year citation format and editable-text requirements.

RESPONSE TO REFEREE 1

1.Novelty of REM and RNI- We clarified that REM is a simplified drainage referenced elevation difference and is conceptually related to HAND; we therefore do not claim that the underlying terrain concept is entirely new. The contribution is framed around its integration and evaluation within the multi-sensor ML framework. RNI is described as a rainfall-conditioned surface-wetness predictor combining cumulative precipitation and NDWI.

2.Threshold justification-The revised Section 4.3 explicitly reports SAR ratio > 1.25 , SAR difference ≥ 3 dB, NDWI > 0.05 and REM < 5 m. The consensus rule is stated explicitly: at least two complementary indicators must agree for a pixel to be labelled flooded. The inventory is described as a consensus-derived reference rather than direct ground truth.

3.Target–predictor dependence-We added an explicit paragraph noting that SAR-derived indicators, NDWI and REM contribute to both the consensus inventory and the predictor set. The revised text explains that this may introduce target–predictor dependence and that the reported metrics therefore represent discrimination against the consensus-derived inventory rather than independent ground-truth accuracy.

4.Independent validation-We agree that independent validation is required. An independent spatially explicit flood inventory for the 2019 event was not available. The historical-location comparison has therefore been renamed “Qualitative Comparison with Documented Flood-Prone Locations” and is explicitly stated not to constitute independent quantitative validation.

5.Statistical significance-The statistical assessment is now described specifically in relation to the evaluated AUC comparisons in the feature-ablation analysis. Claims that could imply overall superiority have been moderated.

6.Cross-validation and robustness-Five-fold cross-validation and sample-size sensitivity analyses are retained and described as stability analyses under the evaluated partitioning and sample-size conditions, rather than as proof of transferability to other flood events or regions.

7.Generalisation and operational claims-We moderated statements concerning generalisation and operational deployment. The framework is described as having potential for future operational applications, subject to independent validation and multi-event testing.

8.Reproducibility-The revised manuscript retains the Zenodo data-availability statement covering processed datasets, hydrologic–topographic predictors, training data and flood labels, together with the stated DOI and CC-BY 4.0 licence.

RESPONSE TO REFEREE 2

1.Methodological quality and clarity-The revised manuscript separates consensus flood-inventory construction from supervised-learning evaluation and presents preprocessing, predictor construction, sampling, model development and evaluation in a more traceable sequence.

2.REM/RNI positioning-REM is now explicitly positioned as a drainage-referenced terrain predictor conceptually related to HAND. RNI is defined according to the implemented formulation as cumulative precipitation combined with NDWI. Unsupported implications that RNI directly measures river distance or connectivity have been removed.

3.Flood-label reliability-The revised text states the thresholds and consensus rule and acknowledges that consensus labels are not equivalent to field observations. The target–predictor dependence limitation is explicitly discussed.

4.Validation-The historical flood-prone-location comparison is presented only as qualitative spatial evidence. The need for independent spatially explicit observations and multi-event validation is stated as future work.

5.Comparison with other studies-Table 10 is retained for contextual comparison, but the manuscript now states that differences in study area, data, label generation, spatial resolution and evaluation procedures prevent direct performance ranking across studies.

6.Urban SAR/optical limitations-The discussion recognises signal ambiguity in complex urban environments and the 30 m spatial-resolution limitation, while avoiding universal claims about alternative methods.

7.Reproducibility-The methods specify the principal model configurations and experiments, and the manuscript retains the deposited processed data and training/label datasets. We avoid claiming public availability of material that is not explicitly deposited.

RESPONSE TO REFEREE 3

1.Research Contribution-We clarified that the contribution is the integration of physically interpretable drainage- and rainfall-conditioned predictors into a multi-source ML workflow and their evaluation through feature-ablation experiments, rather than a claim that every predictor is a fundamentally new hydrologic index.

2.Model performance-The revised manuscript identifies the tuned XGBoost configuration as the selected model within the evaluated experiments and reports 71.7% accuracy and AUC 0.803. Feature-ablation, sample-size and cross-validation results are retained.

3.Hydrologic predictor contribution-The ablation analysis reports the progression from SAR-only to SAR + optical, SAR + optical + hydrologic, and the full framework. Statistical testing is interpreted specifically for the evaluated comparisons.

4.Spatial comparison-Kurla, Saki Naka, Vakola and Dharavi are presented as documented flood-prone locations providing qualitative spatial context. The comparison is explicitly not independent quantitative validation.

5.Limitations and transferability-The revised Discussion and Conclusion identify the single-monsoon scope, consensus labels, lack of an independent spatially explicit inventory and 30 m resolution as limitations, and call for multi-event independent validation.

6.Presentation-Statements that could imply superiority, operational readiness or established transferability have been moderated throughout the Abstract, Discussion, Conclusion and Future Work.

RESPONSE TO REFEREE 4

1.Validation basis-We acknowledge the concern and now make the validation basis explicit: the primary inventory is consensus-derived, while the historical-location comparison is qualitative rather than independent quantitative validation.

2.Potential circularity-We explicitly state that SAR-derived indicators, NDWI and REM contribute both to the consensus inventory and predictor set. The resulting target–predictor dependence limits interpretation as independent ground-truth accuracy.

3.Independent observations-The revised manuscript states that independent spatially explicit observations are required to establish external validity and does not claim such validation where it was unavailable.

4.Performance interpretation-The 71.7% accuracy and 0.803 AUC are presented as results against the consensus-derived inventory under the stated evaluation protocol, not as direct field-observed flood-extent accuracy.

5.Operational applicability-The manuscript now states that the framework has potential to support operational applications, conditional on independent validation across additional flood events and datasets.

RESPONSE TO REFEREE 5

1.Methodological positioning-The revised manuscript presents the workflow from Sentinel-1/2, SRTM, CHIRPS and drainage information through preprocessing, consensus inventory generation, sampling, model training and evaluation more explicitly.

2.Physical interpretation-REM is explained as a drainage-referenced elevation difference conceptually related to HAND. RNI is explained as rainfall-conditioned surface wetness from cumulative precipitation and NDWI. Unsupported claims have been removed.

3.Uncertainty and limitations-The revised Discussion explicitly addresses consensus labels, target–predictor dependence, single-event scope, 30 m resolution and the lack of an independent spatially explicit validation inventory.

4.Literature comparison-The revised text states that differences among published studies prevent direct cross-study performance ranking. Table 10 is therefore used for contextual comparison only.

5.Conclusions-The Conclusion distinguishes the empirical Mumbai findings from future applicability. Transferability and operational use are framed as requiring additional independent validation.

CONSOLIDATED LIST OF RELEVANT MANUSCRIPT CHANGES

1. Clarified the study as an original research contribution and requested correction of the manuscript type in production metadata.
2. Clarified REM as a simplified drainage-referenced elevation difference conceptually related to HAND and moderated novelty claims.
3. Defined RNI according to the implemented formulation using cumulative precipitation and NDWI; removed unsupported river-distance/connectivity implications.
4. Added/clarified flood-inventory thresholds: SAR ratio > 1.25 , SAR difference ≥ 3 dB, NDWI > 0.05 and REM < 5 m.
5. Clarified the consensus rule requiring agreement of at least two complementary indicators.
6. Added explicit discussion of target–predictor dependence arising from shared use of SAR-derived indicators, NDWI and REM.
7. Clarified that reported accuracy/AUC quantify discrimination against the consensus-derived inventory rather than independent ground-truth accuracy.
8. Revised the historical-location analysis to a qualitative comparison and explicitly stated that it is not independent quantitative validation.
9. Retained and clarified five-fold cross-validation and sample-size sensitivity analyses as stability analyses.
10. Clarified statistical assessment of feature-ablation comparisons and moderated interpretation of statistical significance.
11. Moderated claims concerning superiority, generalisation, transferability, robustness and operational readiness.
12. Strengthened limitations concerning the single monsoon cycle, consensus labels, independent validation and 30 m resolution.
13. Retained the literature comparison while stating that cross-study performance ranking is not appropriate because of methodological differences.
14. Corrected the discussion of predictor removal so that reduced performance is described as evidence of contribution within the evaluated experiment, without overstating external validity.
15. Retained/strengthened Data Availability, Author Contributions and Competing Interests declarations required for production.
16. Retained the Zenodo deposition of processed datasets, hydrologic–topographic predictors, training data and flood labels under the stated DOI and CC-BY 4.0 licence.
17. Reviewed citation and presentation consistency, including the required author–year in-text citation format.

FINAL NOTE TO THE EDITOR

We have carefully considered all referee comments and revised the manuscript accordingly. In particular, we have strengthened methodological transparency, clarified the role and limitations of the hydrologic–topographic predictors, explicitly addressed target–predictor dependence, distinguished qualitative spatial comparison from independent validation, and moderated claims that extend beyond the evidence available from the Mumbai case study. We hope that these revisions address the concerns raised during evaluation and improve the scientific clarity and reproducibility of the manuscript.

Authors: Ankush S. Pawar and Gayatri M. Phade