**Review Report**

**Manuscript:** *From Forecast to Alert: Designing an AI-Driven Flood Early Warning System for the White Volta Basin Using Open Satellite Data*

**Reviewer perspective:** Flood Early Warning and Response Systems, GIS/Remote Sensing, information systems architecture, disaster risk management, and people-centred warning practice.

**Overall recommendation:** Major revision before consideration for a high-impact journal.

**Date:** 31 July 2026

---

## 1. Overall assessment

This is a timely and practically useful paper. The authors address a real flood early warning problem in the White Volta Basin, where flood risk in northern Ghana is influenced not only by rainfall but also by water releases from the Bagre Dam in Burkina Faso. This is an important operational issue because downstream communities may experience flood impacts that cannot be explained by local rainfall alone.

The paper has good potential, but it needs **major strengthening**. At present, it is stronger as a technical flood forecasting study than as a complete people-centred flood early warning system paper. The authors should reduce over-strong performance claims, improve validation during severe flood events, and directly compare the proposed model with existing forecasting baselines.

## 2. Main strengths

The paper focuses on a real transboundary flood warning problem. The influence of Bagre Dam is highly relevant for downstream preparedness in Ghana and for other data-scarce river basins affected by upstream reservoir operations.

The study does not only consider discharge prediction. It also attempts to connect AI forecasting, alert thresholds, Sentinel-1 flood mapping and operational dissemination through existing Ghanaian warning arrangements. This is a valuable direction because early warning systems must support decisions and action, not only produce forecasts.

The use of a satellite-derived reservoir storage proxy is the most original contribution of the paper. It could help in situations where formal dam-release information is not available in real time. The use of CHIRPS, ERA5-Land, JRC Global Surface Water, Sentinel-1 SAR and GRDC discharge also shows how open data can help address data scarcity in flood early warning.

## 3. Major points for improvement

### I. Present the model performance more carefully

The paper reports very high model performance. The ensemble model achieved KGE values of 0.984, 0.974 and 0.957 for 1-day, 3-day and 5-day forecasts. These are very strong results. However, the test period does not include the most serious flood events. The authors state that the independent test period does not contain Orange or Red alert events. The paper also states that the GRDC discharge record ends in February 2007, and that the largest historical flood event in 2007 is outside the training, validation and test evaluation.

Therefore, the reported performance values should be interpreted carefully. They show strong performance during the available test period, but they do not fully prove model skill during the most severe flood events. The authors should add event-based validation for peak timing, peak magnitude, threshold crossing, false alarms, missed alerts and duration above threshold.

### II. Show whether AI adds value beyond recent discharge

The paper presents the system as an AI-driven flood early warning system using open satellite data. However, the model appears to depend mainly on recent river discharge. The authors use lagged discharge values and rolling discharge means as input features, including discharge from 1, 2, 3, 5, 7 and 10 days before, plus 3-day and 7-day rolling discharge means.

This is not necessarily wrong, because recent discharge is a useful predictor in flood forecasting. However, the paper claims a strong AI and satellite-data contribution, while the results suggest that much of the model performance may come from recent observed river flow. In simple terms, the model may mainly be learning that if the river was high yesterday and in the last few days, it is likely to remain high tomorrow.

The authors should therefore compare the AI model with simple baselines, including persistence, climatology, linear regression, rainfall-only, Bagre-only and a full model without discharge lags. This would show whether the AI model and satellite-derived variables provide additional value beyond recent discharge.

**III. Validate the Bagre reservoir proxy more strongly**

The Bagre storage proxy is the most important novelty of the paper, so it needs careful validation. The paper should include the exact reservoir boundary, a map of the water mask, sensitivity testing with different polygons, and validation against dam level, reservoir operation data, satellite altimetry or published reservoir information where available.

This is important because the satellite-derived water extent may include surrounding floodplain water, wetlands, river channels or other seasonal water bodies. If this is not checked, the reservoir storage signal may not fully represent the actual Bagre reservoir condition.

**IV. Improve the comparison with GloFAS**

The paper compares the results with a published GloFAS benchmark, but it does not directly compare the proposed model with GloFAS v4 at Nawuni for the same period and using the same metrics. If the authors want to claim operational improvement, they should provide a direct comparison.

A more balanced statement would be that the model performs better than a published GloFAS benchmark, but direct comparison with the current GloFAS product at Nawuni is still needed.

**V. Move from a forecast-alert pipeline to a people-centred FEWRS**

The current architecture is mainly technical: data, model, alert and dissemination. This is useful, but a high-impact early warning paper should also include risk knowledge, exposure and vulnerability, impact-based warning, standard operating procedures, communication pathways, local language messages, community feedback and response capacity.

The use of Log-Pearson Type III distribution to derive discharge thresholds provides a reasonable hydrological basis for the alert levels. However, the paper should go further and explain how these discharge thresholds translate into real-world impacts and response actions. For example, an Orange or Red alert should be linked to affected communities, roads, crops, schools, health facilities, shelters, evacuation routes and institutional responsibilities.

**4. Final recommendation**

The paper has a valuable idea and a strong applied direction, especially through the Bagre Dam storage signal and the attempt to connect forecasting with alert generation. However, I recommend major revision.

The paper should not overclaim severe flood prediction skill without stronger event-based evidence. It should also become more clearly linked to flood early warning and response practice. **With stronger validation, clearer benchmarking and a more people-centred architecture, the paper could make a stronger contribution to both flood forecasting science and disaster risk management practice.**