Journal Homepage: www.ijarpr.com ISSN: 3049-0103 (Online)

International Journal of Advance Research Publication and Reviews

Vol 03, Issue 10, pp 145-150, October, 2026

Disaster Prediction and Early Warning Using Machine Learning

Dr. D. Richard[1], Shivadhanush S[2]

[1]Assistant Professor, Department of Information Technology, St. Joseph's College (Autonomous), Tiruchirappalli, Affiliated to Bharathidasan University, Tiruchirappalli,Tamil Nadu, India, E-mail: richard_it2@mail.sjctni.edu

[2]PG Student, Department of Information Technology, M.Sc. Computer Science Programme, St. Joseph's College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India. Email: shivadhanush5113@gmail.com

Abstract

Disaster early-warning systems increasingly combine environmental sensing, remote sensing, historical records, and machine-learning models.[1] This review-based paper examines how Extreme Gradient Boosting (XGBoost) can support disaster prediction and early warning across flood, landslide, wildfire, and earthquake-related applications. The reviewed literature supports XGBoost most clearly for structured, sensor-driven tasks such as short-horizon flood-stage forecasting, landslide alert-severity prediction, and multi-hazard classification.[2–4] A flood-alert study used five-minute stage forecasts and continuous detection, while a landslide study predicted alert severity ten minutes ahead.[2, 3] A hybrid Neural-XGBoost study reported strong classification performance on a dataset containing wildfires, floods, and earthquakes, but that result should be interpreted as dataset-specific feasibility rather than universal operational validation.[4] The literature also establishes an important boundary for earthquakes: machine learning is more defensible for probabilistic hazard estimation, rapid signal detection, and early warning than for deterministic prediction of an individual event.[5] The paper proposes a hazard-specific, time-aware, leakage-controlled XGBoost workflow that evaluates not only accuracy but also warning lead time, missed events, false alarms, calibration, robustness, and explanation quality. The central conclusion is that XGBoost should be treated as one component of an end-to-end warning system, not as a universal disaster predictor.

Keywords: Disaster prediction; Early warning systems; Machine learning; Extreme Gradient Boosting (XGBoost); Flood forecasting; Landslide monitoring; Wildfire detection; Earthquake early warning; Multi-hazard classification; Explainable artificial intelligence; Sensor-driven prediction; Temporal validation; Warning lead time; Risk assessment.

1. Introduction

India experiences different types of natural disasters because of its large geographical area and variation in climate, terrain and environmental conditions. Floods frequently affect river basins and coastal regions, while cyclones mainly affect coastal states. Earthquakes occur particularly in seismically active regions, and landslides are common in mountainous areas such as the Himalayas and Western Ghats. Droughts affect agricultural regions, while wildfires and extreme heat events can affect forest and dry regions.

Recent events demonstrate the diversity of disaster risks in India. For example, the 2023 Sikkim flash flood was associated with a Glacial Lake Outburst Flood (GLOF) involving South Lhonak Lake and caused major impacts along the Teesta River. Government disaster reports also document recurring flood, heavy-rainfall and landslide incidents across different parts of India.

Disasters can develop rapidly and affect large populations within a short period. Therefore, early identification of dangerous conditions is important for evacuation, emergency response, infrastructure protection and resource allocation.

2. Literature Review

2.1 Machine learning in disaster early-warning systems

A recent systematic review describes AI-driven disaster prediction and early-warning systems as combining machine learning, deep learning, Internet-of-Things sensing, remote sensing, and decision-support functions across natural and human-made hazards.[1] Another review identifies supervised learning, unsupervised learning, ensemble models, deep learning, and explainable AI as major methodological families.[6] This literature indicates that the main research problem is not selecting one universally superior algorithm; it is matching an algorithm and data design to a hazard-specific prediction task.

Early-warning systems generally contain four connected stages: monitoring, analysis, communication, and response.[1] Monitoring produces observations; analysis converts them into a risk estimate; communication translates that estimate into an alert level; and response determines what users should do. A high-performing model that produces a warning too late, cannot tolerate missing sensors, or generates too many false alarms may be less useful than a slightly less accurate model with reliable lead time and clear operating thresholds.

2.2 XGBoost and ensemble learning

XGBoost is used in the reviewed literature as a gradient-boosting learner for outcomes including flood stage, alert severity, and disaster class.[2–4] Its suitability for tabular data makes it attractive when observations come from multiple sensors and when nonlinear relationships and interactions are expected. In contrast, sequence models may be preferable when long temporal dependencies are central and sufficient time-series data are available. The literature therefore supports comparative evaluation rather than assuming that XGBoost is always superior.

2.3 Direct flood and landslide evidence

The most direct evidence concerns floods and landslides. A 2022 flood-alert study used XGBoost to forecast gauge stage levels at five-minute intervals and evaluated a warning system on two historical flooding events in a flood-prone Houston watershed.[2] The study reported agreement between predicted and observed stages, found local gauges useful for estimating flood arrival time in its setting, and described continuous operation that could detect floods without an external event trigger.[2]

For landslides, a field study collected soil-movement data at ten-minute intervals from an actual monitoring site in Mandi, India.[3] Its SVM-XGBoost design used SVM alert predictions as inputs to an XGBoost classifier for ten-minute-ahead alert-severity prediction and compared the ensemble with several baseline and hybrid algorithms.[3] The authors reported that the SVM-XGBoost method outperformed the alternatives for the alerting task.[3] These findings are practically relevant but remain site-specific; a single watershed or slope cannot establish geographic transferability.

2.4 Hybrid, explainable, and multi-hazard approaches

A Neural-XGBoost study combined neural-network feature extraction with XGBoost classification for wildfire, flood, and earthquake classes. It reported 94.8% accuracy and an average F1 score of 0.95 on its dataset, with class-level F1 scores of 0.96 for wildfires, 0.93 for floods, and 0.96 for earthquakes.[4] The result demonstrates a possible hybrid architecture, but performance on one dataset should not be treated as evidence of robust warnings across regions, sensors, or future environmental conditions.

Explainability studies apply LIME and SHAP to inspect model behavior. One study reported 72% accuracy for an XGBoost disaster-classification model and identified features such as “No Injured” and “Total Damages” as influential.[7] Those features may be useful for post-event classification but unavailable before an event; consequently, feature availability must be checked against the intended warning time. Reviews identify interpretability, data quality, class imbalance, and computational complexity as continuing barriers, while recommending hybrid modeling, transfer learning, IoT–AI integration, and explainable AI.[6]

2.5 Forecastability and the earthquake boundary

The literature on earthquake machine learning distinguishes deterministic prediction of an individual earthquake from probabilistic hazard estimation, rapid detection, aftershock forecasting, and early warning.[5] This distinction is essential for responsible terminology. An XGBoost model may estimate whether observed signals are consistent with a hazardous condition or help classify an incoming signal, but that does not mean it can specify the exact time, location, and magnitude of an earthquake before it occurs.

3. Methodology

3.1 Review design and scope

This study uses a narrative, review-based design rather than a new empirical experiment. The evidence was organized into four themes: (1) cross-hazard reviews of machine learning and early-warning systems; (2) direct XGBoost applications in flood and landslide monitoring; (3) hybrid and explainable models; and (4) forecastability, deployment constraints, and responsible claims.[1, 8] The scope is intentionally cross-hazard because XGBoost is a general supervised-learning algorithm used across several disaster tasks.[1, 6] However, the analysis does not assume that hazards share the same labels or forecast horizons. Flood-stage regression, landslide-severity classification, wildfire detection, and earthquake early warning are treated as different tasks. This prevents a high score in one task from being incorrectly presented as evidence for every other hazard. These questions emphasize operational

utility rather than algorithmic novelty. A model was considered particularly relevant when it produced a forecast or alert before, or during, a hazard event and when its output could be connected to a warning decision.

3.3 Conceptual XGBoost workflow

A review-grounded implementation begins by defining the warning target. Examples include water level at a future time, probability of threshold exceedance, landslide-alert severity, or a multi-class hazard type.[2–4] The target must include a time horizon and a label-generation rule. For example, a flood model may predict whether the river will exceed a threshold within the next 30, 60, or 180 minutes, whereas a landslide model may predict alert severity ten minutes ahead.

The second stage is feature construction. Candidate flood features include rainfall intensity and accumulation, river level, flow, reservoir level, soil moisture, temperature, humidity, upstream gauge observations, terrain descriptors, and recent flood labels. Candidate landslide features include rainfall, soil moisture, pore-water pressure, slope movement, terrain slope, and geological descriptors. The reviewed landslide study demonstrates the relevance of soil movement and high-frequency monitoring for short-horizon alerts.[3]

The third stage is temporal alignment. Measurements from sensors should be synchronized to a common clock, and features must be restricted to observations available before the prediction time. This follows the reviewed emphasis on real-time monitoring, data quality, and explainability.[1, 6] Missingness should be recorded explicitly because sensor failure may itself be associated with severe weather. Future values, post-event damage reports, injuries, and labels derived from the outcome should not be used as predictors when the goal is early warning.

3.4 Data preparation and leakage control

The dataset should be divided chronologically rather than randomly. This is a methodological implication of the reviewed studies' site-specific forecast horizons and the literature's concern with generalization and data scarcity.[2, 3, 8] A chronological training period, validation period, and future test period better represent deployment, where the model is trained on past events and then applied to future events. If multiple locations are available, a geographic holdout should also be considered. A model that performs well at one gauge but poorly at unseen gauges has limited evidence of transferability.

Data leakage should be assessed at three levels, consistent with the literature's warning that data quality and evaluation design can limit real-world early-warning validity.[1, 6] First, feature leakage occurs when a predictor contains information measured after the forecast time. Second, event leakage occurs when records from the same disaster appear in both training and test sets. Third, label leakage occurs when the label is constructed from a variable that is also directly supplied as an input. These checks are especially important for disaster data because event records are temporally clustered.

Rare severe events create class imbalance. A model can obtain high overall accuracy by predicting the common non-event class while missing the events that matter most. The literature includes the use of SMOTE in the Neural-XGBoost study.[4] In an operational study, resampling should be performed inside the training folds only, and performance should also be reported on the original event distribution. Class weights, event-based sampling, threshold adjustment, and cost-sensitive learning are alternatives.

3.5 XGBoost model specification

For regression, XGBoost can predict a future water level or continuous hazard index; for classification, it can estimate the probability of an alert class or threshold exceedance. These task types reflect the direct flood, landslide, and multi-hazard applications in the reviewed evidence.[2–4] Important hyperparameters include maximum tree depth, learning rate, number of boosting rounds, row subsampling, column subsampling, minimum child weight, and regularization. Hyperparameter tuning should occur using time-respecting validation rather than random cross-validation.

A useful implementation should compare XGBoost against meaningful baselines. For flood stage, baselines may include persistence, linear regression, random forest, and a hydrological or statistical forecast. For classification, logistic regression, random forest, support-vector machines, and a temporal model such as LSTM may be appropriate when data support them. The SVM-XGBoost landslide study provides a direct example of comparing a hybrid approach with multiple alternatives.[3]

3.6 Operational evaluation

Accuracy alone is insufficient for early warning. The reviewed flood and landslide studies evaluate forecast timing or alert horizon in addition to predictive performance.[2, 3] The evaluation should include recall for severe events, precision, F1 score, area under the precision-recall curve, false-alarm rate, missed-event rate, Brier score, calibration, and warning lead time. A confusion matrix should be reported separately for each warning level. Continuous forecasts should include error measures such as MAE or RMSE, but also threshold-exceedance performance because the operational decision is often categorical.

Warning lead time should be summarized per event rather than only per observation, reflecting the reviewed studies' emphasis on flood arrival time and ten-minute-ahead alerting.[2, 3] The study should report the median and distribution of lead times, the earliest and latest warning, and the fraction of events detected before a minimum action threshold. False alarms should be reported per month, season, or warning opportunity so that emergency managers can interpret alert burden. A model that improves accuracy but warns too late may be less useful than a modestly less accurate model that

consistently provides an actionable window.

Calibration is important because a probability of 0.8 should correspond approximately to an 80% event frequency under comparable conditions. This complements the reviewed literature's emphasis on reliable, interpretable, and operationally useful warnings rather than accuracy alone.[1, 6] Reliability diagrams, Brier scores, and post-hoc calibration can assess this property. Calibration should be checked across seasons, locations, and event severity because an apparently reliable aggregate model may be poorly calibrated for the rare severe events that require intervention.

3.7 Explainability and physical plausibility

SHAP or related feature-attribution methods can identify which observations contribute to an individual warning and which variables dominate across many warnings; the reviewed explainability study demonstrates this type of XGBoost analysis.[7] However, explanation is not the same as causation. Feature importance should be checked for stability across time and sites and interpreted alongside physical knowledge. If the model relies on a variable that is unavailable in real time, the explanation reveals a design flaw rather than a useful scientific insight.

The explainability literature illustrates this issue: variables such as injuries and total damages may contribute to a post-event classification, but they cannot support a warning issued before damage occurs.[7] A responsible methodology therefore includes a feature-availability audit, a leakage audit, and a domain-expert review of the most influential predictors.

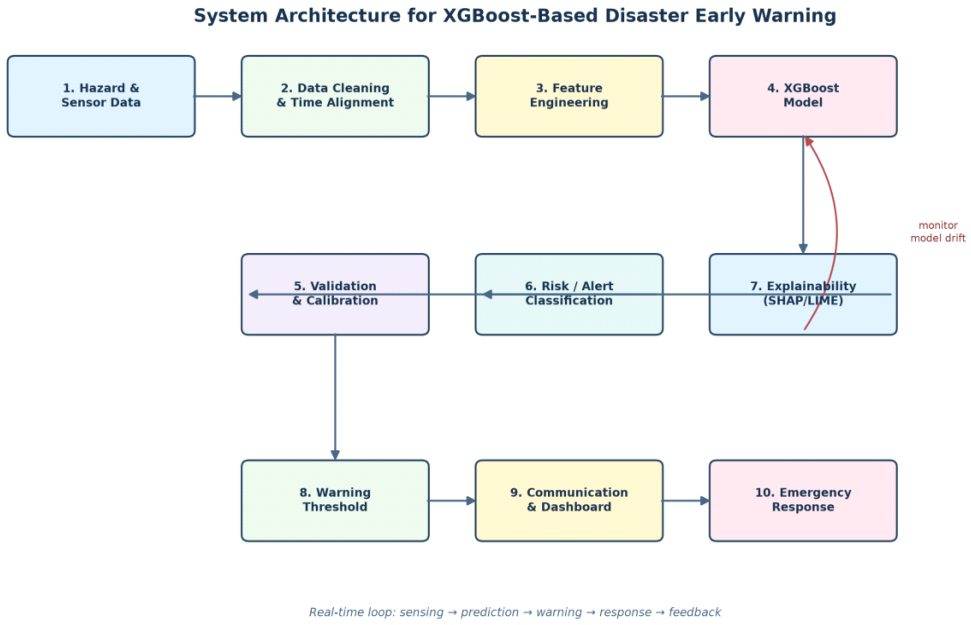
3.8 Robustness, uncertainty, and deployment

Robustness testing should simulate missing sensors, delayed data, noisy measurements, seasonal shifts, and changes in event frequency. If the model is intended for multiple regions, external geographic validation is essential. Transfer learning or domain adaptation may help data-scarce regions, but successful transfer must be demonstrated rather than assumed; reviews identify data scarcity and generalization as continuing challenges.[6, 8]

Uncertainty should be communicated to the operator. This recommendation follows the broader literature's emphasis on explainable, deployable, and decision-support-oriented warning systems.[1, 6] Prediction intervals for continuous stage forecasts, calibrated probabilities for classifications, and an "insufficient data" state for severe sensor failure are preferable to a forced binary answer. The model should be monitored after deployment for drift, changes in sensor behavior, changes in land use, and changing climate conditions. Finally, the model must be connected to an alert protocol. A warning level should have a defined threshold, message, recipient, response action, escalation rule, and cancellation rule. Human review may be appropriate for high-consequence alerts. This system-level design follows the literature's emphasis on integrating AI with sensing, communication, explainability, and emergency management rather than treating the model as a stand-alone predictor.[1, 6]

4.System Architecture Workflow Diagram

The proposed architecture connects real-time hazard observations to preprocessing, XGBoost prediction, validation, explainability, warning thresholds, communication, and emergency response. The feedback path represents monitoring of model drift and operational performance after deployment.



5. Results and Discussion

5.1 Synthesis of reviewed evidence

The strongest direct evidence supports XGBoost for structured, high-frequency monitoring tasks. In flood forecasting, XGBoost was used for five-minute stage predictions and continuous flood detection in a specific watershed.[2] In landslide monitoring, an SVM-XGBoost ensemble predicted alert severity ten minutes ahead and outperformed the compared alternatives in the reported site study.[3] These studies suggest that XGBoost is especially useful when sensor observations are frequent, the target is clearly defined, and the warning horizon is short enough for recent measurements to contain predictive information.

The hybrid Neural-XGBoost study broadens the possible use case to multi-hazard classification. It reported 94.8% accuracy and an average F1 score of 0.95, with class-level F1 scores of 0.96 for wildfires, 0.93 for floods, and 0.96 for earthquakes.[4] These numbers are promising, but they do not establish operational reliability because the result is tied to one dataset, its sampling design, its labels, and its train-test procedure. The appropriate interpretation is that hybrid representation learning plus XGBoost is a plausible architecture requiring external validation.

5.2 Interpretation for practice

The evidence favors a hazard-specific development strategy. Flood and landslide models can begin with XGBoost because the literature provides direct examples of short-horizon forecasting and severity alerting. A multi-hazard model should be introduced only after individual hazard labels, sensor pipelines, and evaluation procedures are validated. Earthquake applications require especially careful wording: the literature supports probabilistic estimation, rapid detection, aftershock forecasting, and early warning, not deterministic prediction of the exact future event.[5] The principal research gap is not another isolated accuracy comparison. It is evidence of transferability, calibration, reliable lead time, resilience to missing data, and human usefulness under realistic deployment conditions. Published studies often focus on a single site or dataset, while reviews continue to identify data quality, class imbalance, interpretability, and computational complexity as barriers.[6, 8] The next generation of studies should therefore report event-level and geographic holdout results and connect model outputs to explicit emergency actions.

5.3 Limitations of this review

This paper is a narrative synthesis rather than a formal meta-analysis, and the reviewed studies use different hazards, datasets, outcomes, metrics, and validation designs.[1, 8] Direct numerical comparison across studies is therefore not justified. The conclusions are strongest where multiple sources converge on the same methodological point—such as the importance of operational timing, explainability, and hazard-specific validation—and weaker where evidence comes from a single application.

6. Conclusion

The reviewed evidence supports XGBoost as a useful component of disaster early-warning systems, particularly for structured sensor-driven tasks such as flood-stage forecasting, landslide alert-severity prediction, and selected multi-hazard classification problems.[2–4] Its responsible use requires a clearly defined hazard, forecast horizon, leakage-controlled temporal validation, imbalance-aware training, operational metrics, calibrated probabilities, and explanations based only on information available at warning time. The evidence does not support presenting XGBoost as a universal deterministic disaster predictor, especially for individual earthquake occurrence.[5] The most defensible future research is a hazard-specific, externally validated, end-to-end warning evaluation that measures whether predictions arrive early enough and reliably enough to support action.

7. References

- **AI-Driven Disaster Prediction and Early Warning Systems: A Systematic Literature Review.** 2024. Review of machine learning, deep learning, IoT, remote sensing, and early-warning applications across hazards.
- **Early Natural Disaster Prediction Using Machine Learning: A Comprehensive Review.** 2025. Review of supervised, unsupervised, deep-learning, ensemble, IoT, and explainable-AI approaches.
- **Application of Machine Learning in Early Warning System of Geotechnical Disaster: A Systematic and Comprehensive Review.** 2025. Systematic review of machine learning for geotechnical-disaster prediction and warning.
- **Data-Driven Flood Alert System (FAS) Using Extreme Gradient Boosting (XGBoost) to Forecast Flood Stages.** 2022. XGBoost flood-stage forecasting at five-minute intervals and continuous alert generation.
- **Development of an Ensemble Gradient Boosting Algorithm for Generating Alerts About Impending Soil Movements.** 2021. SVM-XGBoost ensemble for ten-minute-ahead landslide alert-severity prediction.

- **Neural-XGBoost: A Hybrid Approach for Disaster Prediction and Management Using Machine Learning.** 2025. Neural feature extraction combined with XGBoost classification for wildfire, flood, and earthquake classes.

- **Utilizing Explainable AI Methodologies: LIME and SHAP, for the Classification of Natural Disasters Through Machine Learning Algorithms.** 2024. Explainability and XGBoost-based disaster classification.

- **Machine Learning for Earthquake Prediction: Techniques for Seismic Forecasting and Disaster Preparedness.** 2024. Distinction between deterministic earthquake prediction, probabilistic forecasting, rapid detection, and early warning.

- **Disaster Prediction Using Appropriate Machine Learning Techniques.** 2024. Overview of machine-learning techniques used for natural-disaster prediction.

- **Machine Learning in Emergency Management: Analysis and Methods.** 2024. Machine-learning applications spanning forecasting, early detection, warning, monitoring, damage assessment, and response.