

Flood Disaster Prediction and Early Warning System Using Machine Learning

A Case Study Using Hourly Environmental and Hydrological Data from the Godavari Basin

Payal Vasanta Khandare

Sadhu Vaswani Institute of Management Studies for Girls, Pune,

Abstract: Floods are difficult to predict because rainfall, river or water level, soil moisture, weather conditions, and their recent changes interact over time. A practical early warning system therefore needs to provide predictions early enough for people and authorities to respond. This study presents flood disaster prediction and early warning system developed using hourly environmental and hydrological data from the Godavari Basin. The dataset contains 8,760 hourly observations and 25 columns, including rainfall, water level, soil moisture, temperature, humidity, atmospheric pressure, wind conditions, cumulative rainfall features, and lagged water-level and soil-moisture features. The proposed system uses 20 non-leaking input features and predicts three risk horizons: 1 hour, 3 hours, and 6 hours. The target classes are Normal, Alert, and Flood. Data were cleaned, timestamps were converted to a consistent time format, missing input values were handled using median imputation, and chronological train-test evaluation was used to reduce information leakage from future observations. Multiple machine-learning models were compared, including Extra Trees, XGBoost, Logistic Regression, and Random Forest. XGBoost was selected for the multi-horizon implementation because it provided a strong balance between accuracy and class-sensitive performance and was trained with class balancing to reduce the effect of the minority Flood class. For the reported 1-hour XGBoost test result, the confusion matrix contained 1,752 test observations and produced 96.06% accuracy, 87.40% macro precision, 86.58% macro recall, and 86.98% macro F1-score. Flood-class precision, recall, and F1-score were 69.70%, 66.99%, and 68.32%, respectively. The final system provides a simple Gradio-based interface that accepts environmental conditions and returns 1-hour, 3-hour, and 6-hour predictions together with class probabilities and an overall warning message. The study demonstrates that a lightweight, data-driven multi-horizon approach can support flood early warning while remaining simple enough for academic and prototype deployment.

Keywords: Flood prediction; early warning system; machine learning; XGBoost; Godavari Basin; rainfall; water level; time-series classification; multi-horizon forecasting; disaster management.

I. INTRODUCTION

Floods are among the most damaging natural hazards and can cause loss of life, infrastructure damage, agricultural losses, and disruption of economic activity. Accurate and timely flood forecasting is therefore an important part of disaster risk reduction. Traditional flood forecasting approaches can be highly effective, but they may require detailed physical, hydraulic, and hydrological information and can be computationally demanding. Machine learning (ML) provides a complementary data-driven approach because it can learn complex relationships between historical environmental observations and flood-related outcomes [1], [4], [5].

Recent research shows a wide range of ML and deep-learning applications for flood forecasting, monitoring, susceptibility mapping, and early warning. Reviews have reported strong use of rainfall, water level, streamflow, and related meteorological variables, while ensemble methods, hybrid models, and deep-learning methods continue to grow [1], [4], [5]. At the same time, practical flood warning systems must address several challenges, including class



imbalance, missing data, changing hydrological conditions, uncertainty, model interpretability, and the difference between predicting a flood event and delivering a useful warning to an end user.

The present research focuses on a practical classification-based early warning problem. Instead of producing only one forecast, the proposed system produces predictions for three future horizons: 1 hour, 3 hours, and 6 hours. This multi-horizon design is intended to provide a more useful view of how risk may develop over time. The model uses current environmental conditions together with short-term historical and lagged features. The final predictions are integrated into a simple software interface so that a user can enter environmental conditions and receive a risk class and probability-based warning.

The main contributions of this study are: (1) development of a three-class flood-risk classification framework for Normal, Alert, and Flood conditions; (2) use of 20 environmental and lag-based features without the current-risk label as a model input; (3) comparison of several machine-learning algorithms; (4) use of class balancing to improve attention to the minority Flood class; (5) evaluation using a chronological train-test strategy; and (6) development of a simple multi-horizon early-warning interface for 1-hour, 3-hour, and 6-hour predictions.

II. LITERATURE REVIEW

Machine learning has been widely studied for flood prediction because it can model nonlinear relationships among rainfall, water level, streamflow, soil conditions, and meteorological variables. Mosavi, Ozturk, and Chau reviewed a large body of flood-prediction literature and reported that ensemble methods, hybridization, optimization, and data decomposition are important directions for improving prediction quality [1]. More recent reviews continue to show that ML and deep learning are being applied to short-term forecasting, flood susceptibility, flood mapping, and early-warning systems [4], [5], [6].

Computer-vision approaches are another important research direction. Dhaya and Kanthavel developed a video-surveillance flood monitoring system in which water-level information was extracted from images and classified using a convolutional neural network [2]. Their work shows that visual monitoring can complement sensor-based forecasting, although image-based systems can require suitable camera placement and robust performance under different environmental conditions.

Hybrid hydrological and deep-learning models have also been investigated. Nguyen et al. combined HEC-HMS with an encoder-decoder LSTM for real-time flood forecasting and reported improvements at longer forecast horizons [3]. This suggests that combining physical hydrological knowledge with data-driven learning can be valuable, but such systems may require more domain information and calibration than a lightweight classification system.

Recent work further emphasizes real-time and explainable warning. A 2025 review comparing short-term flood forecasting studies found that the best-performing approaches vary with forecast horizon and that RF, LSTM, ANN, ANFIS, CNN, and hybrid methods can each be effective under different conditions [5]. In 2026, Potharaju et al. presented an explainable deep unsupervised approach for flood monitoring using multivariate water-level sensors, highlighting the importance of anomaly detection and interpretability [7]. Wiangnak et al. developed an AIoT-based early warning framework combining visual water-level detection and LSTM forecasting, demonstrating the value of integrating sensing and forecasting for real-time warning [8].

These studies demonstrate strong progress, but several practical gaps remain. First, many studies focus on one prediction horizon or on continuous water-level forecasting rather than a simple multi-class risk warning. Second, advanced deep-learning and multi-source systems may require larger datasets, more computational resources, and additional sensors. Third, class imbalance is often important when severe flood events are relatively rare. Finally, a prediction model is more useful when its outputs are converted into a clear warning that a non-specialist user can understand. The present study addresses these gaps by combining multi-horizon classification, class balancing, chronological evaluation, and a simple user-facing warning interface.



2.1 Selected Literature and Research Gap

Study	Year	Approach	Main Finding	Gap Relevant to This Study
Mosavi et al. [1]	2018	ML literature review	ML methods can provide effective and cost-efficient flood prediction; ensemble and hybrid methods are important.	More work is needed for practical short-term warning and broader applicability.
Dhaya & Kanthavel [2]	2022	CNN + image features	Video surveillance can support automatic water-level monitoring.	Image systems need broader environmental validation and are not a substitute for time-series forecasting.
Nguyen et al. [3]	2024	HEC-HMS + Encoder-Decoder LSTM	Hybrid modelling improved real-time flood forecasting, especially at longer horizons.	Requires watershed-specific hydrological modelling and broader validation.
Li & Jun [4]	2024	Review of ML for river flood routing	ML can capture nonlinear flood-routing relationships with less physical-process specification.	Model reliability depends on data quality and appropriate validation.
Asif et al. [5]	2025	Review/intercomparison of short-term ML forecasting	Performance varies with forecast horizon; RF, LSTM, ANN, ANFIS, CNN and hybrid methods are all relevant.	Need practical systems that balance performance, complexity and usability.
Potharaju et al. [7]	2026	Explainable LSTM autoencoder	Anomaly detection can identify unusual hydrological behaviour and provide explanations.	Needs larger and geographically diverse validation.
Wiangnak et al. [8]	2026	AIoT + YOLOv10 + LSTM	Combining visual sensing and time-series forecasting can support real-time early warning.	More validation across regions and infrastructure conditions is needed.

2.2 RESEARCH GAP

Based on the reviewed literature, the research gap addressed in this work is the need for a simple and practical multi-horizon flood-risk classifier that uses commonly available environmental and historical variables, explicitly considers the minority Flood class, and converts predictions into an understandable warning. Rather than depending on images, IoT hardware, or a large number of external data sources, the proposed framework uses a compact set of 20 tabular features. It also predicts three future horizons so that the user can see whether risk remains normal or becomes more concerning over the next several hours.

III. METHODOLOGY

The proposed methodology consists of six stages: data preparation, target definition, feature preparation, chronological model training, model evaluation, and deployment through a prediction interface. The overall workflow is shown below.



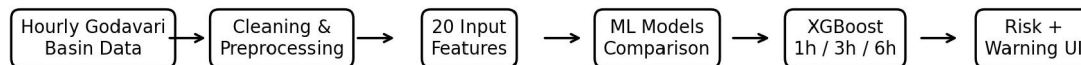


Figure 1. Overall workflow of the proposed multi-horizon flood prediction and early-warning system.

3.1 Dataset Description

The study uses an hourly Godavari Basin dataset containing 8,760 observations and 25 columns. The variables represent rainfall, water level, soil moisture, temperature, humidity, pressure, wind conditions, cumulative rainfall over different periods, lagged water-level observations, lagged soil-moisture observations, and three future risk targets. The timestamp covers approximately June 2024 to May 2025. The three target variables are target_1h, target_3h, and target_6h, representing risk at 1-hour, 3-hour, and 6-hour horizons.

The input set was restricted to 20 environmental and historical features. The current_risk field was excluded from the model inputs because it can contain information directly related to the risk label and may create target leakage. The selected features were rainfall_mm, water_level_m, soil_moisture, temperature_c, humidity_pct, pressure_hpa, wind_speed_ms, wind_direction_deg, rainfall_1h_mm, rainfall_3h_mm, rainfall_6h_mm, rainfall_12h_mm, rainfall_24h_mm, water_level_lag_1h_m, water_level_lag_3h_m, water_level_lag_6h_m, water_level_lag_12h_m, water_level_lag_24h_m, soil_moisture_lag_6h, and soil_moisture_lag_12h.

3.2 Target Class Distribution

The 1-hour target is imbalanced, with Normal observations substantially more common than Flood observations. Before removing the single missing 1-hour target, the distribution was 5,835 Normal, 2,411 Alert, 513 Flood, and 1 missing value. After removing the missing target, 8,754 observations remained for the 1-hour task. This imbalance is important because a model can obtain high overall accuracy while still missing a meaningful proportion of rare Flood cases.

Target Horizon	Normal	Alert	Flood
1 hour	5,830	2,411	513
3 hours	5,788	2,162	804
6 hours	5,749	2,018	987

3.3 Data Preprocessing

The timestamp field was converted to a datetime representation and the observations were ordered chronologically. Missing values in the input features were handled using median imputation. Invalid target observations were removed separately for each horizon. This approach avoids using future information to fill the target and provides a stable preprocessing procedure for deployment.

Because the data are hourly time-series observations, a chronological train-test split was used rather than a random split. The reported XGBoost 1-hour confusion matrix contains 1,752 test observations, which corresponds to a chronological 80/20 evaluation of the 8,754 valid 1-hour records. This evaluation strategy is more realistic for an early-warning system because the model is trained on earlier observations and evaluated on later observations.

3.4 Feature Engineering

The model combines current conditions and historical context. Current variables describe the environmental state at the prediction time, while cumulative rainfall features represent rainfall over recent time windows. Lagged water-level and soil-moisture variables provide temporal information that may capture gradual hydrological changes. These features allow the classifier to learn relationships between recent environmental conditions and future risk.

- Current: rainfall, water level, soil moisture, temperature, humidity, pressure, wind speed, wind direction.
- Rainfall history: 1-hour, 3-hour, 6-hour, 12-hour and 24-hour rainfall.



- Water-level history: 1-hour, 3-hour, 6-hour, 12-hour and 24-hour lags.
- Soil-moisture history: 6-hour and 12-hour lags.

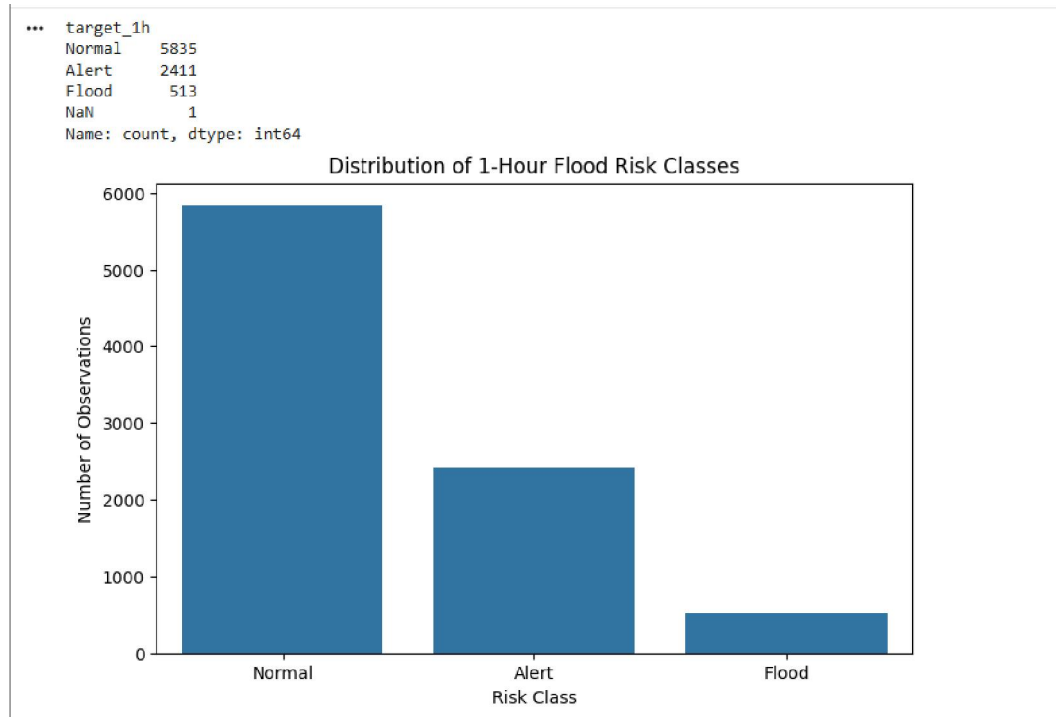


Figure 2. Distribution of the 1-hour flood-risk classes in the study dataset

3.5 Machine Learning Models

Four classification approaches were compared in the modelling experiments: Extra Trees, XGBoost, Logistic Regression, and Random Forest. Tree-based ensemble methods are useful for this problem because the relationship between environmental variables and flood risk can be nonlinear. Logistic Regression provides a simple linear baseline, while Extra Trees and Random Forest provide ensemble-based nonlinear baselines. XGBoost was selected for the final multi-horizon implementation because it provides strong predictive performance while remaining practical for tabular data.

To address class imbalance, balanced sample weights were used during model training. This is especially important because the Flood class is much smaller than the Normal class in the 1-hour target. The final implementation trains separate XGBoost models for target_1h, target_3h, and target_6h, rather than forcing one model to represent all forecast horizons.

3.6 Evaluation Metrics

The models were evaluated using accuracy, precision, recall, and F1-score. Accuracy measures the proportion of all correct predictions. Precision measures how many predictions assigned to a class are correct. Recall measures how many actual observations of a class are detected. F1-score is the harmonic mean of precision and recall. Macro-averaged metrics give equal importance to each class and are therefore useful when the target distribution is imbalanced.

For the confusion matrix, accuracy is calculated as:



$$Accuracy = (TP + TN + \dots) / Total\ observations$$

For each class, precision = $TP / (TP + FP)$, recall = $TP / (TP + FN)$, and $F1 = 2 \times Precision \times Recall / (Precision + Recall)$.

IV. RESULTS AND DISCUSSION

The study produced a complete classification and early-warning pipeline. The class distribution confirms that flood events are relatively rare compared with normal conditions, while the monthly pattern shows that the flood class is concentrated in the earlier part of the study period. This concentration makes chronological evaluation particularly important because a random split could place very similar time periods in both training and test sets and may give an overly optimistic estimate of future performance.

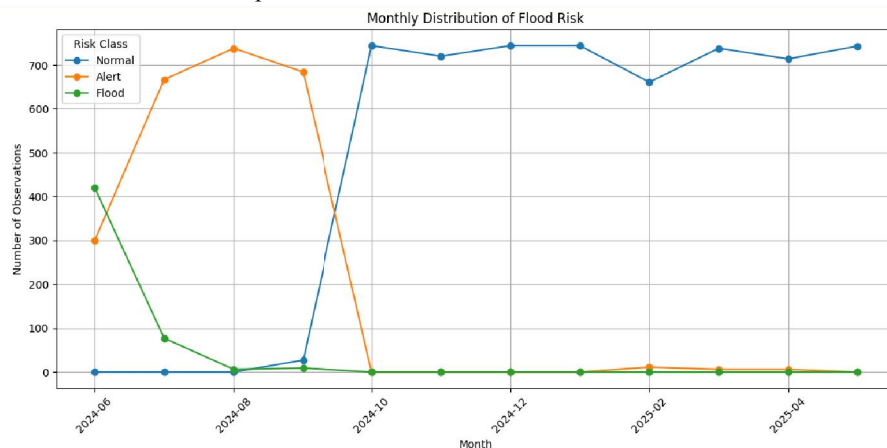


Figure 3. Monthly distribution of Normal, Alert, and Flood observations.

Figure 3 shows a clear temporal change in the target distribution. Flood observations are concentrated around June 2024, while Normal observations dominate most later months. This pattern suggests that the dataset contains a strong seasonal and event-related component. A model that performs well on this dataset should therefore be evaluated on future time periods rather than only on randomly selected observations.

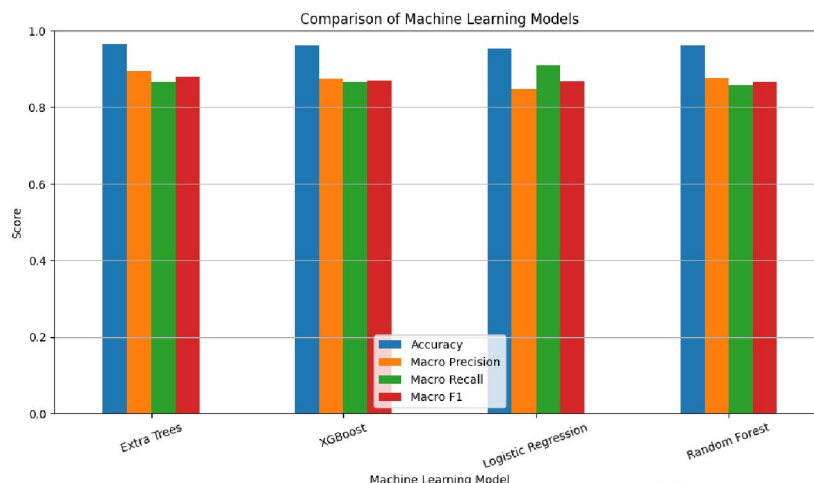


Figure 4. Comparison of the evaluated machine-learning models using accuracy, macro precision, macro recall, and macro F1-score.



The model-comparison plot shows that all four evaluated models achieved high overall accuracy, while their macro precision, macro recall, and macro F1-score were lower than their accuracy. This difference is expected in an imbalanced three-class problem because the majority Normal class can contribute strongly to accuracy. The comparison supports the use of macro-averaged metrics and class-sensitive evaluation instead of relying on accuracy alone.

4.1 XGBoost 1-Hour Test Performance

The XGBoost 1-hour confusion matrix contains 1,752 test observations. The diagonal entries are 447 Alert, 69 Flood, and 1,167 Normal predictions. The resulting accuracy is 96.06%. The macro precision is 87.40%, macro recall is 86.58%, and macro F1-score is 86.98%.

Metric	Alert	Flood	Normal
Precision	92.93%	69.70%	99.57%
Recall	92.74%	66.99%	100.00%
F1-score	92.83%	68.32%	99.79%

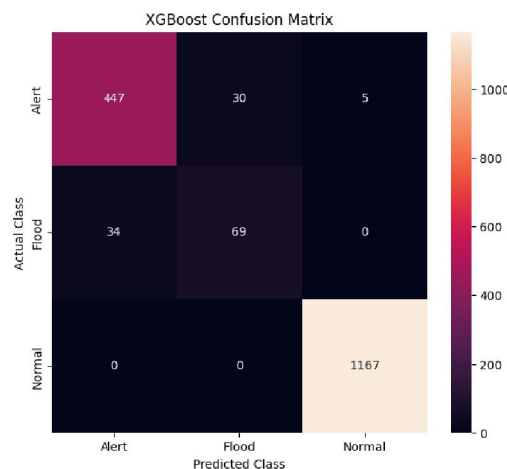


Figure 5. XGBoost confusion matrix for the 1-hour test set.

The confusion matrix shows that the model identifies the Normal class very strongly, with all 1,167 Normal test observations correctly classified. For the Flood class, 69 observations were correctly identified as Flood and 34 were classified as Alert, with no Flood observation classified as Normal. This is an important result for an early-warning application because the most serious error would be to classify an actual Flood case as Normal. However, the Flood recall of 66.99% also shows that the system still misses a meaningful number of flood cases by predicting Alert instead. Therefore, the model should be considered a decision-support prototype rather than a replacement for official hydrological warnings.

The results also demonstrate why class balancing and macro metrics are important. The Normal class has a much larger number of observations, so a model could achieve high accuracy even with weak Flood performance. In the reported XGBoost result, the Flood class achieved 69.70% precision and 66.99% recall, showing that the model has learned useful information about severe events but still has room for improvement.

4.2 Multi-Horizon Prediction and Early Warning Interface

The final system uses three separately trained models. Given one set of 20 environmental and historical inputs, the system generates three predictions: 1-hour, 3-hour, and 6-hour. For each horizon, the interface reports the predicted



class and the estimated probabilities for Normal, Alert, and Flood. The three outputs are then combined into an overall warning message.

The warning logic is intentionally simple. If any horizon predicts Flood, the system reports a flood-related overall status, with the maximum Flood probability used to distinguish between higher and lower warning levels. If no horizon predicts Flood but at least one predicts Alert, the overall status is Alert/Moderate Risk. If all horizons predict Normal, the system reports Normal/Low Risk unless the estimated Flood probability remains elevated. This design converts model probabilities into information that is easier for a user to interpret.

The user interface is implemented with Gradio and can run in Google Colab. The software design separates model loading, preprocessing, prediction, and presentation. The trained models and supporting objects can be saved to Google Drive and loaded again when the Colab runtime is restarted. This makes the prototype easier to reproduce without retraining the models every time the notebook is reopened.

4.3 Discussion

The experimental findings are consistent with the broader literature in one important respect: tree-based and ensemble models can perform strongly on structured hydrological and meteorological data [1], [4], [5]. The current study also supports the observation that the best model depends on the prediction task, dataset, forecast horizon, and input variables. Unlike image-based flood monitoring [2] or sensor-intensive AIoT systems [8], the proposed method focuses on tabular environmental data and can therefore operate as a lightweight software prototype.

A major strength of the proposed approach is its multi-horizon design. A single flood classification at one future time may not provide enough information about how risk is developing. Predictions at 1, 3, and 6 hours allow a user to compare near-term and slightly longer-term risk. However, the horizons are produced by separate predictive models and should not be interpreted as independent official forecasts. Further validation with real operational warning records, additional locations, and longer time periods is required before practical deployment.

Another important observation is the difference between overall accuracy and Flood recall. The 96.06% accuracy is encouraging, but the Flood recall of 66.99% is more relevant to warning performance. In disaster applications, false negatives can be more serious than false positives. Future work should therefore investigate threshold tuning, cost-sensitive learning, calibration, uncertainty estimation, and possibly hybrid or deep-learning models. Recent literature also emphasizes explainability, spatial information, and integration of multiple data sources [5], [7], [8].

V. LIMITATIONS

- The dataset covers approximately one year, so the model may not represent all seasonal and extreme-event patterns.
- The flood class is smaller than the Normal class, creating class imbalance and making Flood recall a critical concern.
- The current evaluation is based on one study region and should not be assumed to generalize to other river basins without validation.
- The system uses tabular environmental inputs and does not currently integrate satellite imagery, radar, traffic, topographic maps, or real-time IoT streams.
- The current interface requires manual input of the 20 variables. A future operational version should automatically receive validated sensor or API data.
- Model probabilities are predictive scores and should not be interpreted as official emergency thresholds without calibration and domain validation.
- The system is a research prototype and should complement, not replace, official meteorological and hydrological warning systems.



VI. FUTURE WORK

Future work can improve the proposed system in several directions. First, the dataset can be expanded to multiple years and additional river-basin locations. Second, satellite rainfall, radar, topography, land-use, and real-time sensor data can be integrated to improve spatial awareness. Third, LSTM, Temporal Convolutional Networks, Transformers, hybrid hydrological-ML models, and ensemble stacking can be compared with the current XGBoost approach. Fourth, explainable AI methods such as SHAP can be used to show which environmental features contributed most to each prediction. Fifth, probability calibration and uncertainty estimates can be added so that the system communicates not only a class but also the confidence and reliability of the prediction. Finally, the Gradio prototype can be converted into a web or mobile service connected to live environmental data and validated with official warning records.

VII. CONCLUSION

This study developed an AI-based multi-horizon flood disaster prediction and early warning system using hourly environmental and hydrological data from the Godavari Basin. The proposed framework uses 20 non-leaking features and predicts Normal, Alert, and Flood conditions at 1-hour, 3-hour, and 6-hour horizons. The data were preprocessed using chronological ordering and median imputation, while class balancing was used to reduce the impact of the imbalanced target distribution. Four machine-learning approaches were compared, and XGBoost was used for the final multi-horizon implementation.

For the reported 1-hour XGBoost test set, the model achieved 96.06% accuracy, 87.40% macro precision, 86.58% macro recall, and 86.98% macro F1-score. The Flood class achieved 69.70% precision, 66.99% recall, and 68.32% F1-score. These results indicate that the model can identify useful patterns in environmental and historical observations, while also showing the need for further improvement in detecting all severe flood cases. The final Gradio interface converts the predictions into simple multi-horizon risk information and warning messages. Overall, the work demonstrates the feasibility of combining machine learning and a lightweight user interface for research-oriented flood early warning. Before real-world use, the model must be independently validated on longer and geographically diverse datasets and aligned with official hydrological warning procedures.

REFERENCES

- [1] A. Mosavi, P. Ozturk, and K.-W. Chau, "Flood Prediction Using Machine Learning Models: Literature Review," *Water*, vol. 10, no. 11, 1536, 2018. doi: 10.3390/w10111536.
- [2] R. Dhaya and R. Kanthavel, "Video Surveillance-Based Urban Flood Monitoring System Using a Convolutional Neural Network," *Intelligent Automation & Soft Computing*, vol. 32, no. 1, pp. 183–192, 2022. doi: 10.32604/iasc.2022.021538.
- [3] P. S. Nguyen, T. H. (Felix) Nguyen, and T. Nguyen, "A real-time flood forecasting hybrid machine learning hydrological model for Krong H'nang hydropower reservoir," *River*, vol. 3, no. 1, pp. 107–117, 2024. doi: 10.1002/rvr2.72.
- [4] L. Li and K. S. Jun, "Review of Machine Learning Methods for River Flood Routing," *Water*, vol. 16, no. 2, 364, 2024. doi: 10.3390/w16020364.
- [5] M. Asif, M. M. Kuglitsch, I. Pelivan, et al., "Review and Intercomparison of Machine Learning Applications for Short-term Flood Forecasting," *Water Resources Management*, vol. 39, pp. 1971–1991, 2025. doi: 10.1007/s11269-025-04093-x.
- [6] V. Kumar, H. M. Azamathulla, K. V. Sharma, and D. J. Mehta, "The State of the Art in Deep Learning Applications, Challenges, and Future Prospects: A Comprehensive Review of Flood Forecasting and Management," *Sustainability*, vol. 15, no. 13, 10543, 2023.
- [7] S. Potharaju, S. N. Tambe, R. K. Tirandasu, et al., "Explainable deep unsupervised learning for flood monitoring and anomaly detection using multivariate DWLR sensors," *Scientific Reports*, vol. 16, 25377, 2026. doi: 10.1038/s41598-026-55978-9.



- [8] V. Wiangnak, M. Wiboonrat, and S. Duangsuwan, "Development of an AIoT-Based Early Flash-Flood Warning System for Smart Rural Disaster Resilience," *Sensors*, vol. 26, no. 11, 3512, 2026. doi: 10.3390/s26113512.
- [9] "Flood susceptibility mapping: Integrating machine learning and GIS for enhanced risk assessment," *Applied Computing and Geosciences*, vol. 23, 100183, 2024. doi: 10.1016/j.acags.2024.100183.
- [10] J. L. R. Neypes, M. Sultana, and A. J. R. Miyajima, "Real-time Flood Surveillance and Early Warning System using IoT-AI Fusion for Flood Prediction: Smart Framework to Monitoring Solutions," *International Journal of Multidisciplinary and Current Educational Research*, vol. 6, no. 4, pp. 30–40, 2024.

DATA AND FIGURE NOTE

The numerical results, class distributions, confusion matrix, and model-comparison figures in this manuscript are based on the author's current Godavari Basin project outputs. Figures 2–5 were supplied from the project notebook. The paper should be updated if the final trained models, dataset version, or evaluation split changes. Before international submission, the author should also add the exact dataset source/collection description, institutional affiliation, supervisor/co-author details, ethics or data-use statement if required, and the target journal's formatting and reference style.

