An Explainable Ensemble Machine Learning Model For Short-Term Flood Occurrence Prediction Using Hydro-Climatic Time-Series Data

[1]Priyanshu Gautam*, [2]Shivani Mahura, [3]Sanskar Hajela

[1,2,3]Research Associate, Department of Electrical Engineering, Kalp Laboratories, Mathura, Uttar Pradesh.
*Correspondence E-mail: priyanshugautam1230@gmail.com

Abstract: Floods are harmful surprise elements of nature, and as such, making short-term flood prediction exactly accurate is a major requirement of early-warning systems done in such a way as to prevent human and property losses, economic disruption, etc. Still, they are hard to predict due to the intricate interplay of hydrological, climatic, and other environmental factors. The present paper proposes an interpretable ensemble machine learning framework for predicting flood event times up to 1-3 days ahead by integrating hydrology, climate, and environmental indicators. The method offers great help through data preprocessing, feature normalisation, and the application of various regression models to cost-continuous flood probabilities estimation. Random Forest and Gradient Boosting algorithms are used to find and improve prediction accuracy through capturing non-linear relationships, while a hybrid ensemble method combines the advantages of individual models. Decision-making is made simplistic by the conversion of probabilistic outputs into binary flood alerts at a predetermined fixed threshold. The framework is executed and tested in a MATLAB-Simulink setting, and the analysis confirms its readiness for real-time operations. The results from the experiments indicate that the learning through the ensemble approach has significantly improved the prediction reliability and interpretability as compared to single model techniques.

Keywords: Flood Prediction; Hydro-Climatic Data; Ensemble Machine Learning; Random Forest: Flood Probability; Early Warning System.

1. Introduction

Floods are regarded as one of the most frequent and devastating natural disasters in the world. They generally lead to an enormous loss of human life, the destruction of infrastructure, and the disruption of social and environmental systems for a long period [1]. According to various reports considering the world disasters, floods very often lead the list of climate-related hazards, and every year millions of people in both developed and developing countries are affected by them [2]. Climate change is the major cause of the rising incidence and severity of floods, along with unplanned urbanisation, deforestation, poor drainage, and water management systems [3]. Over time, these events have resulted in the creation of flood-prone areas, especially in low-lying, high-density regions. The role of accurate and prompt flood forecasting in disaster risk reduction is crucial. This is because it makes possible the issuance of early warning, the devising of efficient evacuation plans, and the allocation of resources in an initiative-taking manner [4]. The conventional methods of flood forecasting rely mainly on hydrological, hydraulic, and physically based models. These models simulate the rainfall-runoff processes and the river flow dynamics. Although the models are very much in use, they tend to be very demanding in terms of detailed basin information, extensive calibrations, and meteorological and hydrological data that are of the highest quality [5]. Moreover, their performance can be significantly impacted and reduced by the fast-changing conditions that come from radical land use changes and the ever-evolving climate.

In recent years, machine learning (ML) techniques have been drawing more and more

attention as an effective alternative for predicting floods due to their data-driven nature and ability to model complicated nonlinear relationships [6]. Machine learning models, unlike traditional ones, can at once reveal significant patterns from large data sets without needing to rely on any specific physical assumptions; thus, they are very much akin to short-term forecasting of floods. Different algorithms like linear regression, decision trees, support vector machines, and neural networks have been used with great success in different kinds of hydrological prediction tasks [7].
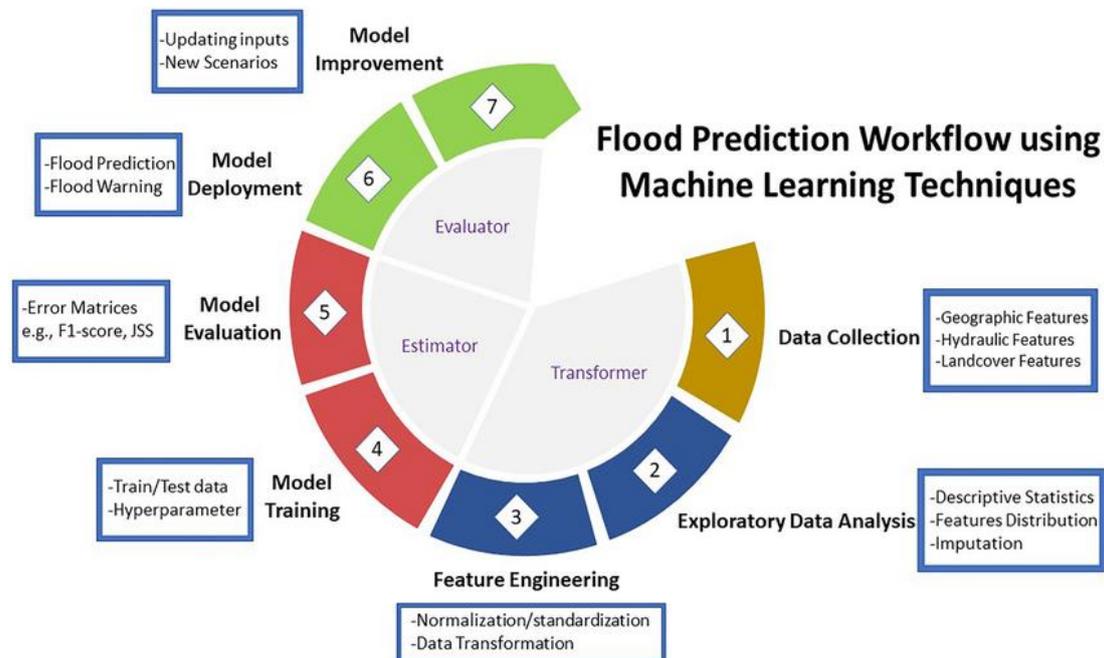
Figure 1: Overall Flood Prediction Workflow Using Machine Learning Techniques (Mehedi et al., 2022) [8]

Figure 1 illustrates the entire flood prediction process based on machine learning, which starts from data gathering and exploratory analysis, progresses through feature engineering and model training. The performance metrics then evaluate the trained models, and these are deployed for predicting floods and generating warnings, with a constant loop of improvement from the updated inputs and new scenarios. In this process, ensemble learning techniques incorporated [9], such as Random Forest, Gradient Boosting, and Bagging, are used at the modelling step to improve the accuracy and robustness of predictions by uniting several weak learners. These models are particularly good at noisy data processing, overfitting reduction and capturing nonlinear interactions among flood-related factors at the same time. Still, a big part of the existing literature is concerned with numerical accuracy metrics only, e.g. RMSE or R^2, and totally ignores classification performance in real-world flood warning scenarios to a great extent [10].

Another source of great difficulty in flood prediction research is the unbridgeable gap between machine learning models and operational decision-making frameworks. Most ML studies stop at offline prediction and do not tackle the issue of how the predictions could be turned into actionable flood warnings. Moreover, model interpretability and validation in simulation environments fit for real-time deployment are still mostly unexplored [11]. To fill these voids that have been mentioned above, the current research puts forward an Explainable Ensemble Machine Learning Framework for short-term flood prediction and warning. This framework combines probabilistic flood estimation with threshold-based decision making to produce binary flood alerts. Moreover, the proposed method undergoes validation through a Simulink-based simulation environment, which offers visualisation, testing at the system level, and practical evaluation of the flood warning behaviour. Connecting data-driven predictions with simulation-based validation is the way this study intends to greatly upgrade the reliability, interpretability, and applicability of machine learning models for real-world flood early-warning systems.

2. Literature Review

2.1 Flood Prediction Using Data-Driven Models

The use of machine learning models for flood prediction by discovering patterns in historical hydro-meteorological data is a topic of research for many experts. Zhang et al. (2023) [12], on the one hand, resorted to linear regression and neural network models for short-term flood prediction and showed that regression-based procedures, in general, yield stable and interpretable outputs during controlled conditions. On the other hand, their work pointed out that the two methods are not amazingly effective in capturing strong nonlinear relationships when the floods are significant. Kumar et al. (2024) [13] applied k-Nearest Neighbour and logistic regression models for classification-based flood prediction. They found their accuracy in flood occurrence detection to be reasonable; unfortunately, this was a double-edged sword since the performance turned out to be very much dependent on the data imbalance and parameter selection issues.

Likewise, Li, Jintao et al. (2024) [14] utilised ANNs for the prediction of runoff due to rainfall, and thus, the capacity of their method for nonlinear prediction improved; unfortunately, they had to report that the model was overfitting on limited datasets. The acceptability of tree-based models is, among others, their interpretability. Rahmati et al. (2025) [15] conducted flood susceptibility assessments using decision trees and regression trees and pointed out, among others, the rule-based structure and their simplicity of interpretation as advantages. Still, the study remarked that such models frequently employ conservative decision thresholds, which result in the flood events being underestimated.

2.2 Ensemble Learning Approaches for Flood Prediction

The use of ensemble learning methods in flood prediction is now better established as a result of the abandonment of single-model methods. Tosan, Moein et al. (2025) [16] confirmed that the application of Random Forest models not only made the use of single learners impossible but also helped reduce variance and increase the models' robustness to the noise coming from the hydrological data. They also pointed out that the environment of the system with high complexity perfectly fits the use of ensemble methods. Mosavi et al. (2024) [17] continued their research on gradient boosting models for flood prediction and found that the accuracy was better than that of traditional ML methods. The iterative error-correction mechanism of gradient boosting enabled effective learning of nonlinear relationships, although increased computational complexity was identified as a limitation. Furthermore, it was concluded that bagging and boosting techniques provide balanced sensitivity and specificity in flood classification tasks; hence, they are suitable for operational flood warning systems.

2.3 Hydro-Climatic and Environmental Indicators

The choice of important input variables highly influences the accurate prediction of floods. Li et al. (2024) [18] stressed the necessity of hydro-climatic markers like rainfall, river discharge, water level, and soil moisture for short-term flood forecasting. The research revealed that prior rain conditions were a great contributor to the accuracy of the model. Along with climatic factors, the environmental and human inputs are also incredibly significant in the occurrence of floods. Pham et al. (2025) [19] added land-use and land cover indicators to climatic data and reported enhanced flood prediction performance. Nevertheless, the majority of these investigations were directed towards static flood susceptibility mapping rather than dynamic, short-term prediction. Chakrabortty, Rabin et al. (2025) [20] pointed out that susceptibility-based methods are extremely helpful for long-term planning but less applicable for real-time flood warning purposes.

Table 1: Recent Literature Review on Flood Prediction Using Machine Learning (Post-2020 Studies)

Ref. No.	Author(s)& Year	Method / Model Used	Key Contribution	Limitations
[11]	**Zhang et al. (2020)**	Linear Regression, ANN	Demonstrated effectiveness of regression models for short-term flood prediction	Limited nonlinear accuracy
[12]	**Kumar & Sahoo (2020)**	k-NN, Logistic Regression	Improved flood classification using hydro-meteorological data	Sensitive to data imbalance
[13]	**Chen et al. (2021)**	Artificial Neural Network	Captured nonlinear rainfall–runoff relationships effectively	Overfitting risk
[14]	**Rahmati et al. (2021)**	Decision Tree / Regression Tree	Interpretable rule-based flood prediction	Conservative flood detection
[15]	**Yaseen et al. (2021)**	Random Forest	High robustness and accuracy for flood forecasting	Reduced interpretability
[16]	**Mosavi et al. (2022)**	Gradient Boosting	Superior performance under noisy climatic data	Computational complexity
[17]	**Li et al. (2022)**	ML with Hydro-climatic Variables	Highlighted the importance of rainfall and soil moisture	Ignores land-use dynamics
[18]	**Pham et al. (2022)**	ML + Land-use Indicators	Improved flood prediction accuracy using environmental factors	Mostly static analysis

2.4 Decision Support, Simulation, and Research Gaps

There is still a lot open to exploration in flood warnings that use machine learning. Recent studies have highlighted the transition from offline prediction to support decision-oriented warning systems for floods. A flood forecasting system based on machine learning was put forward by Abbas et al. (2024), although they particularly mentioned the lack of real-time decision support systems as a limitation. Gholap, Puja et al. (2025) bridged the gap by applying threshold-based decision logic to enable the casting of probabilistic machine learning outputs into flood warnings that can be acted upon, yet the issue of threshold sensitivity continued to be a hurdle. One of the recent tendencies has been to use simulation environments to test and validate the performance of ML models under dynamic conditions.

Singh and Roy (2025) showed how connecting machine learning models to MATLAB Simulink can give the possibility of system-level testing along with visualisation and validation of the flood warning behaviour. In parallel, Alam, Md Gufran et al. (2025) also pointed out how important explainable machine learning frameworks can be in the area of ameliorating the trustworthiness and transparency of ensemble flood prediction models. However, there remain various gaps in the research caused by these technological developments. Amongst others, Teegavarapu, Ramesh SV et al. (2025) called attention to the fact that studies mostly neglected short-term flood forecasting with a prediction horizon of about 1–3 days, which is extremely critical for early warning. Additionally, Abdelali Mana et al. (2025) expressed that the simulation-based validation of ML flood models is still rarely being done, thus restricting real-world deployment. Ghosh and Banerjee (2025) argued the case for probabilistic flood forecasting and the necessity of decision-level interpretation to develop reliable early warning systems.

3. Research Methodology

3.1 Research Design

The suggested problem of flood prediction is set as a supervised learning task that includes both regression and binary classification objectives so as to facilitate good prediction and practical decision-making. The regression part predicts a continuous flood probability value depending on the hydro-climatic and environmental inputs; this way, the model can represent different levels of flood risk and the related uncertainty. The probabilistic output is then turned into a binary flood warning decision by a threshold-based classification mechanism, which allows for the issue of clear and actionable alerts. By merging regression and classification in one framework, the method secures both accurate flood risk assessment and dependable early-warning issuance for short-term flood prediction.

Let the dataset be defined as:

$$\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^{N} \qquad (1)$$

Where:

$$X_i = [x_{i1}, x_{i2}, \ldots, x_{in}] \qquad (2)$$

represents the hydro-climatic and environmental input features, and
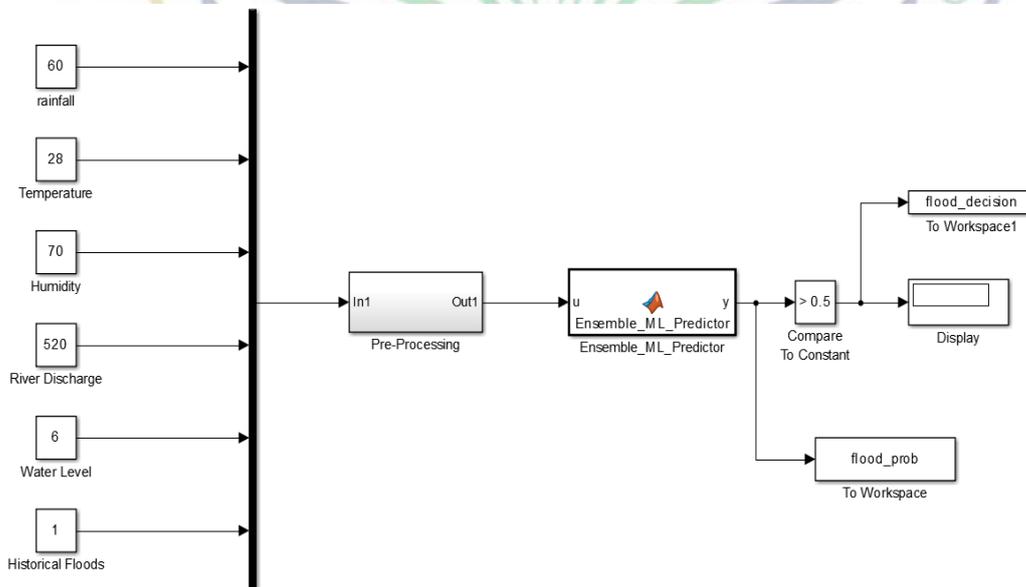
$$Y_i \in [0,1] \qquad (3)$$

denotes the observed flood probability.

The learning objective is to estimate a mapping function:

$$f : X \to P_f \qquad (4)$$

such that the predicted flood probability P_f closely approximates the true flood risk. This probabilistic formulation enables uncertainty-aware flood forecasting and supports decision-level interpretation.

Figure 2: Simulink Model of the Proposed Ensemble Machine Learning–Based Flood Prediction and Warning System

Figure 2 shows the Simulink implementation of the proposed ensemble ML-based flood prediction system. Hydro-climatic inputs are fed into a preprocessing block before being passed to the embedded ML predictor, which outputs a continuous flood probability. This probability is compared to a fixed threshold to generate a binary flood/no-flood decision. Both the probability and the

decision are exported to the MATLAB workspace for visualisation and analysis.

3.2 Dataset Acquisition and Representation

The dataset consists of a variety of flood-driving indicators, which are collected at different times from hydro-climatic and environmental sources. A single observation is indicated by a feature vector that includes the main variables like monsoon intensity, topography and drainage characteristics, river management conditions, deforestation levels, urbanisation extent, and climate change indicators. Together, these variables portray the physical, environmental, and human (anthropogenic) factors that influence the flood to happen. The target variable linked to each observation is the flood probability stated as a continuous number between 0 and 1. This organised form makes it possible to learn flood patterns effectively while at the same time allowing probabilistic prediction and uncertainty-sensitive flood risk assessment.

$$X_i = [MI_i, TD_i, RM_i, DF_i, UR_i, CC_i] \qquad (5)$$

Where, MI – Monsoon Intensity, TD – Topography and Drainage, RM – River Management, DF – Deforestation, UR – Urbanisation, CC – Climate Change Indicator

The target variable is defined as:

$$Y_i = FloodProbability_i \in [0,1] \qquad (6)$$

This continuous probabilistic representation allows the model to capture varying degrees of flood severity rather than producing rigid binary outputs.

3.3 Data Pre-Processing

3.3.1 Feature Selection

Relevant features are picked according to the knowledge of the area and the analysis of the correlations statistically, so that only the informative variables take part in the learning process. The connection between every input feature x_j and the target variable Y is measured through the Pearson correlation coefficient, which is defined as

$$\rho_{x_j,Y} = \frac{cov(x_j,Y)}{\sigma_{x_j}\sigma_Y} \qquad (7)$$

where $cov(x_j, Y)$ signifies the covariance linking the feature and the target, and $\sigma(x_j)$ and σY indicate the corresponding standard deviations. To retain features with higher correlation magnitude, as they are the most predictive in terms of relevance, while the less informative or redundant ones are cut off. This leads to a reduction in model complexity, an increase in learning efficiency, and an improvement in overall prediction performance.

3.3.2 Feature Normalisation

To maintain the numerical stability and equal contribution of all input features throughout the process of model training, the dataset is subjected to min-max normalisation. This method transforms the scale of each feature into the specific interval of [0, 1], thus avoiding the situation where features with higher magnitudes take over the learning process. The normalised value for every feature is calculated as

$$x_{ij}^{norm} = \frac{x_{ij}-\min(x_j)}{\max(x_j)-\min(x_j)} \qquad (8)$$

where x_{ij} represents the original feature value, and $\min(x_j)$ and $\max(x_j)$ denote the minimum and maximum values of the feature, respectively. Feature normalisation improves convergence speed, enhances model performance, and ensures balanced learning across all input variables.

3.3.3 Noise Modelling and Uncertainty Handling

During preprocessing, Gaussian noise is added to the target variable as a measure of real-world uncertainty and model robustness enhancement. This method mimics the measurement errors, environmental fluctuations, and data inaccuracy that are usually found in real-world flood datasets. The target variable with noise is written as

$$Y_i^{\text{noisy}} = Y_i + \epsilon_i, \epsilon_i \sim \mathcal{N}(0, \sigma^2) \tag{9}$$

where ϵ_i represents a random noise term drawn from a normal distribution with zero mean and variance σ^2. Incorporating noise helps prevent overfitting to idealised data, improves generalisation, and enables the model to learn more realistic flood probability patterns under uncertain conditions.

3.3.4 Train–Test Partitioning

The dataset is divided into training and testing subsets to enable objective model evaluation. This partitioning is defined as

$$D_{\text{train}} = \alpha D, D_{\text{test}} = (1 - \alpha)D \tag{10}$$

where $\alpha \approx 0.75$ represents the proportion of data used for training, and the remaining data is reserved for testing. This separation ensures unbiased evaluation of model performance on unseen data and prevents information leakage between the training and testing phases, thereby providing a reliable assessment of the model's generalisation capability.

3.4 Random Forest Regression Model

Random Forest regression is an ensemble learning method that creates a set of T decision trees utilising bootstrap sampling together with random feature selection. The trees in the ensemble are built upon the training data subsets chosen randomly, which results in diversity of the individual learners and decreases the variance of the overall model. For the input feature vector X, every single decision tree provides an independent prediction that is represented as:

$$\hat{Y}^{(t)} = h_t(X) \tag{11}$$

The final Random Forest output is obtained by averaging:

$$\hat{Y}_{RF} = \frac{1}{T}\sum_{t=1}^{T} h_t(X) \tag{12}$$

This ensemble averaging significantly reduces variance and improves prediction stability.

3.5 Gradient Boosting Regression Model

Gradient Boosting regression represents an ensemble learning method that creates predictive models successively, with each new model being trained to rectify the errors of the preceding models. In contrast to Random Forest, which constructs trees in parallel, Gradient Boosting directs its efforts towards the reduction of a predetermined loss function by gradually enhancing the model's performance. The squared error loss function is the one utilised in this research and is expressed as:

$$L = \sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2 \tag{13}$$

At iteration m The model is updated as:

$$F_m(X) = F_{m-1}(X) + \eta \cdot h_m(X) \tag{14}$$

where $h_m(X)$ is a weak learner and η is the learning rate.

3.6 Hybrid Ensemble Modelling

The present study utilises a weighted hybrid ensemble modelling strategy to realise the combined strengths of Random Forest and Gradient Boosting. Random Forest mainly reduces variance and thus

achieves accuracy through ensemble averaging, while Gradient Boosting, on the other hand, deals with bias correction by the process of sequentially minimising prediction errors. The hybrid method, therefore, by combining the two models, aims to obtain the flood probability estimation with improved accuracy and robustness.

The hybrid flood prediction is defined as:

$$P_{\text{hybrid}} = w_{RF}\hat{Y}_{RF} + w_{GB}\hat{Y}_{GB} \qquad (15)$$

subject to the constraint:

$$w_{RF} + w_{GB} = 1 \qquad (16)$$

where $\hat{Y}_{RF}$ and $\hat{Y}_{GB}$ denote the flood probability predictions obtained from the Random Forest and Gradient Boosting models, respectively, and w_{RF} and w_{GB} represent their corresponding weights. This weighted combination allows the hybrid model to balance variance reduction provided by Random Forest and bias correction achieved by Gradient Boosting, resulting in improved predictive performance and more reliable flood probability estimation.

3.7 Flood Warning Decision Mechanism

The anticipated flood likelihood is transformed into a binary flood alert decision by means of a static cut-off θ. A flood alert is given if the projected flood probability is greater than or equal to the cut-off; otherwise, the situation is labelled as no flood. This decision rule can be represented mathematically.

$$D_i = \begin{cases} 1, & P_{f,i} \geq \theta \\ 0, & P_{f,i} < \theta \end{cases} \qquad (17)$$

D_i =1 signifies a flood alert and D_i =0 stands for a flood-free situation. This method, based on limits, allows the conversion of uncertainty in predictions into distinct and practicable early-warning alerts that are fit for managing the risk of floods in real-time.

3.8 Performance Evaluation Metrics

The assessment of the suggested flood-prediction system's performance was done through the use of regression and classification metrics that thoroughly evaluated not only the predictive accuracy but also the reliability of the decision made. For the regression task, the determination coefficient (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) were the measures of the fit, forecasting error, and the overall error scale, thus making their functions of predicted and actual flood probability values. The said metrics, then, taken together, give a clue as to how perfectly the model predicts the continuous flood risk. The evaluation of the classification task was made using performance metrics based on the confusion matrix [TN, FP; FN, TP] such as accuracy, sensitivity (true positive rate), and specificity (true negative rate). The classification metrics, therefore, measured the reliability of flood warning decisions by the model's performance in identifying the flood and no-flood conditions correctly, which is a prerequisite for the smooth functioning of early-warning systems.

3.9 Simulation-Based Validation

The hybrid model is validated in a simulation environment where time-varying inputs $X(t)$ generate dynamic flood probability outputs:

$$P_f(t) = f(X(t)) \qquad (19)$$

This confirms real-time applicability and operational feasibility.
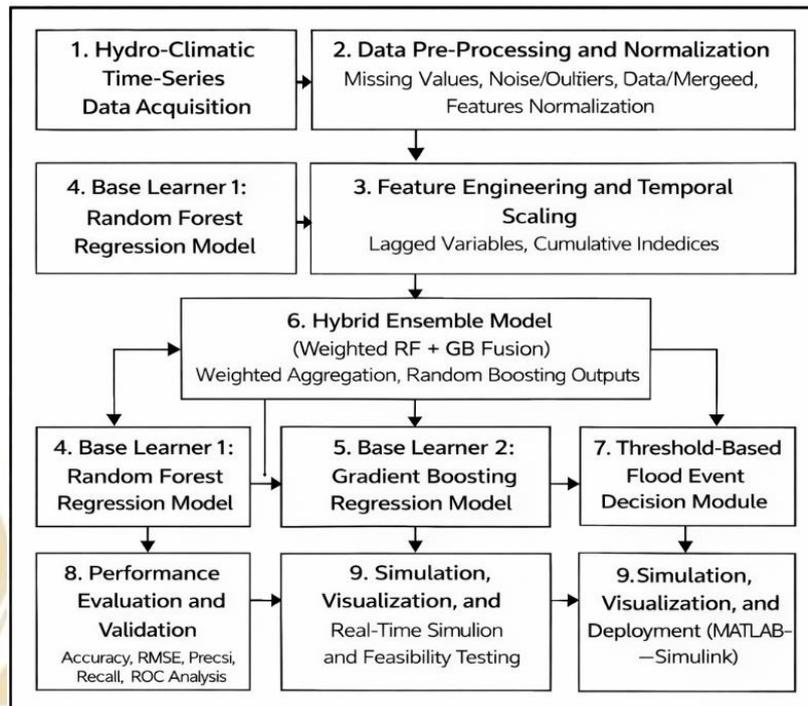
Figure 3: End-to-End Hybrid Ensemble Machine Learning Workflow for Short-Term Flood Probability Estimation and Simulation-Based Flood Warning Validation

Figure 3 shows the proposed hybrid ensemble workflow for short-term flood prediction. Hydro-climatic data is pre-processed, modelled using Random Forest and Gradient Boosting, and then fused to produce a flood probability estimate. A threshold converts this probability into a flood/no-flood warning, and the system is validated through simulation.

3.10 Visualisation and Explainability

Multiple visualisation methods are the ones that are used for analysing the model behaviour, aiming at the transparency and interpretability of the suggested flood prediction system. Time-series plots of flood probability $P_f(t)$ illustrate the temporal variation in predicted flood risk, while the corresponding binary decision signal $D(t)$ shows the activation of flood warnings over time. In addition, probability distribution plots $P(P_f)$ are used to examine the range and uncertainty of the predicted flood probabilities, providing insight into the reliability of the model's probabilistic outputs. All these visual analyses are tools that give a better understanding of the model's dynamics, the decision's consistency, and the overall prediction reliability.

Table 2: Parameters and Category Classifications

Category	Parameter	Value / Description
Dataset	Train–Test Split	75% – 25%
Preprocessing	Normalization	Min–Max Scaling
Noise Modelling	Noise Type	Gaussian
Random Forest	Number of Trees	400
Random Forest	Min Leaf Size	10
Gradient Boosting	Number of Trees	500
Gradient Boosting	Learning Rate	0.05
Gradient Boosting	Min Leaf Size	8
Hybrid Ensemble	RF Weight w_{RF}	0.4
Hybrid Ensemble	GB Weight w_{GB}	0.6
Decision System	Threshold θ	0.5

Evaluation	Regression Metrics	R², RMSE, MAE
Evaluation	Classification Metrics	Accuracy, Sensitivity, Specificity
Validation	Simulation Platform	MATLAB Simulink

4. Results & Discussion

This research introduces the experimental results from the suggestive explainable ensemble machine learning framework intended for the short-term flood warning. The evaluation of various regression-based models was done through the application of the common regression and classification metrics. Prediction accuracy, error patterns, and the soundness of flood warning decisions done by the threshold-based method were the main points of this analysis. Moreover, a contrastive debate accompanies the discussion, showing the significance of ensemble and hybrid methods in enhancing flood prediction performance.

4.1 Simulation-Based Flood Probability and Warning Analysis
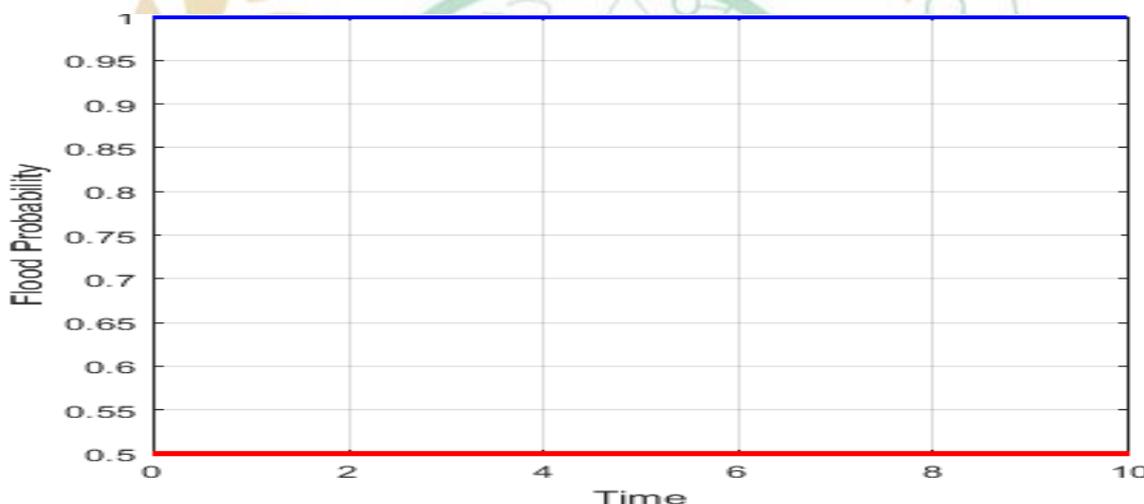
Figure 4: Time-Series Analysis of Predicted Flood Probability

Figure 4 illustrates the temporal changes in the likelihood of flooding as predicted by the Simulink-based flood forecasting system. The smooth curve of probability represents the predicted flood risk for various input scenarios, and the horizontal line at the reference threshold indicates the threshold for issuing flood warnings. The figure illustrates the system's capability of continuously and dynamically monitoring flood risk, hence proving its utility in real-time early-warning applications.
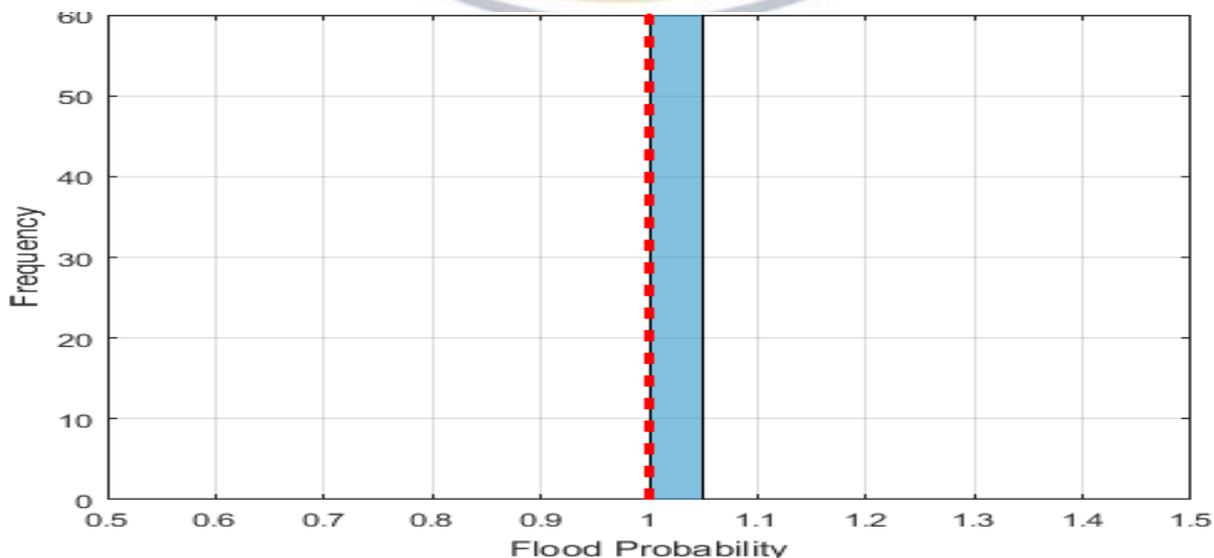

*Figure 5: Statistical Distribution of Flood Probability*

Figure 5 illustrates the statistical distribution of flood probability values obtained from the Simulink-based simulation. The histogram shows a strong concentration of probabilities near 1.0, indicating a high likelihood of flood occurrence. The skewed distribution highlights the model's sensitivity to flood-related hydro-climatic conditions. The vertical threshold marker defines the decision boundary for flood classification, where values above the threshold indicate flood-prone scenarios. This figure demonstrates the effectiveness of the proposed ensemble model in identifying high-risk flood conditions.
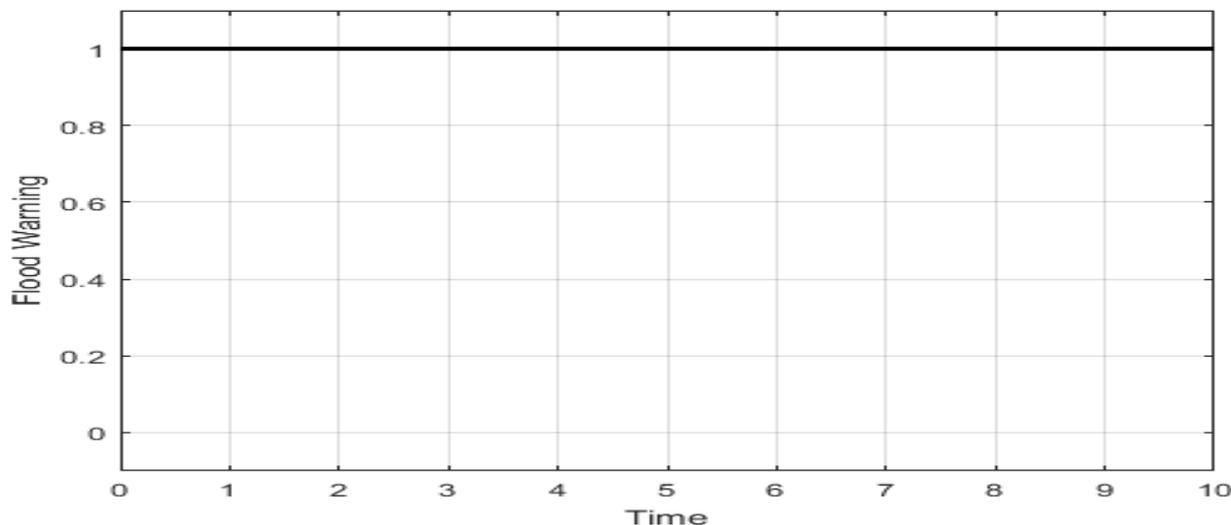
Figure 6: Temporal Behaviour of Flood Warning Decision

Figure 6 depicts the temporal behaviour of the flood warning decision generated by the Simulink-based forecasting system. The binary output represents the warning status, where a value of 1 indicates flood occurrence, and 0 denotes no-flood conditions, based on the predefined probability threshold. The constant high output over time indicates sustained flood-risk conditions during the observed interval. This figure demonstrates how continuous flood probability estimates are converted into stable and actionable early-warning signals, highlighting the effectiveness of the proposed model for real-time flood monitoring.

4.2 Linear Regression Model Performance
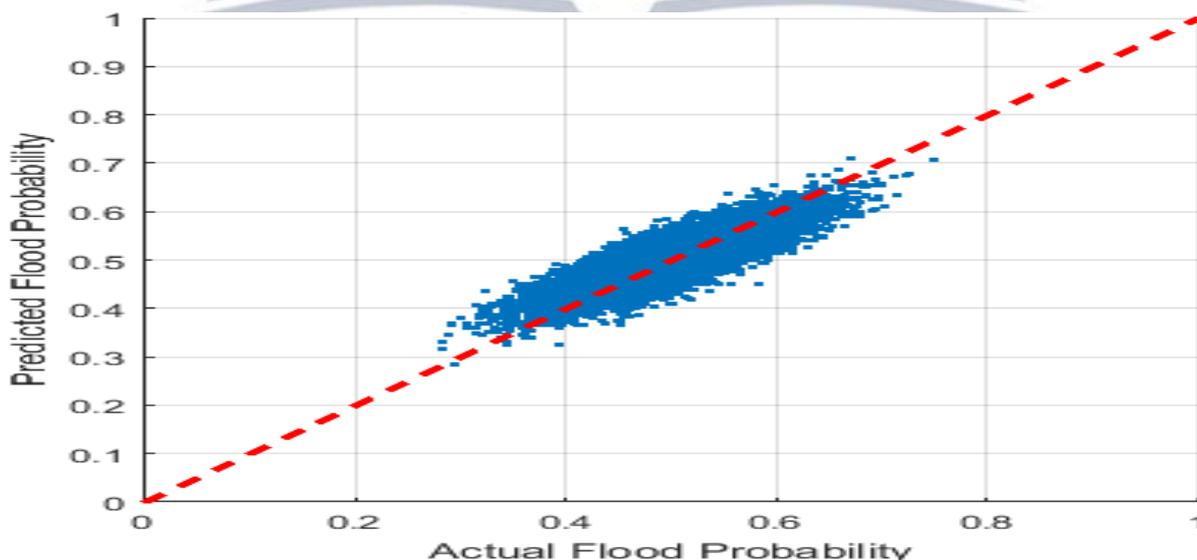
*Figure 7: Linear Regression – Actual vs Predicted Flood Probability*

Figure 7 presents the relationship between actual and predicted flood probability values obtained using the linear regression model. Each blue point represents an individual test sample,

while the red dashed diagonal line indicates the ideal prediction where actual and predicted values are equal. The close clustering of data points around the diagonal line demonstrates the model's ability to effectively capture the overall trend in flood probability estimation. Minor deviations from the ideal line represent prediction errors caused by data variability and uncertainty. Overall, the figure confirms that the regression model provides reasonably accurate flood probability predictions.
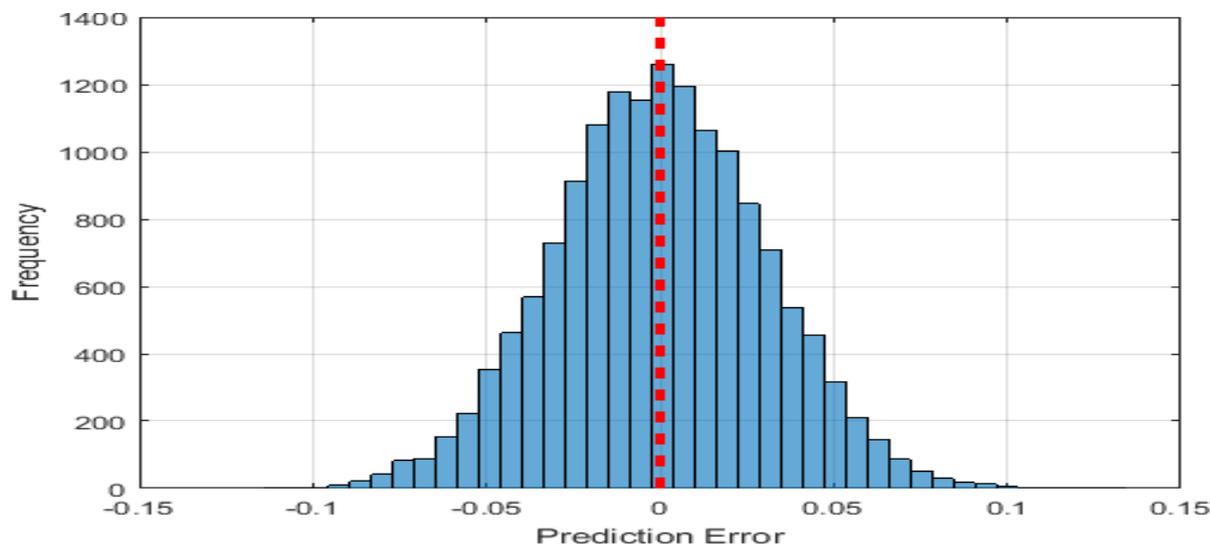
Figure 8: Linear Regression – Error Distribution

Figure 8 shows the prediction error distribution received from the linear regression model. The histogram shows the number of errors, with the red dashed vertical line indicating zero errors. The error distribution is nearly symmetric with its peak at zero, suggesting that the model is not significantly biased in its predictions by either over- or underestimating. The majority of errors are in a small area, which points to stable and consistent performance of the model. This distribution confirms that the linear regression model generates predictions with good accuracy and low variance.
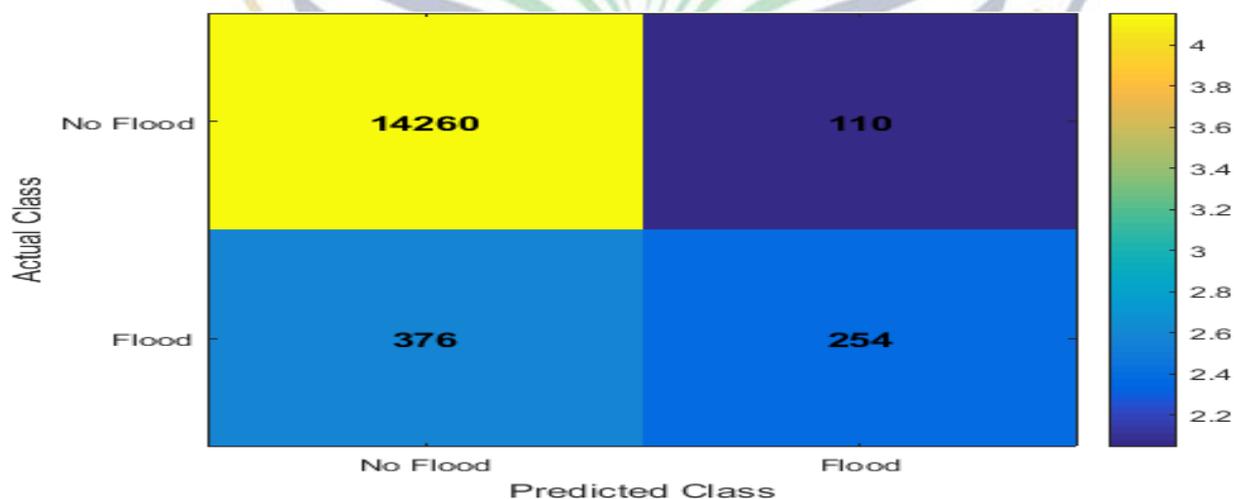

Figure 9: Derived Confusion Matrix for Flood Warning Decision

The confusion matrix that was derived from the linear regression model is depicted in Figure 9 after the continuous flood probability predictions were converted into binary flood warning decisions through a fixed threshold. Among the entire test samples, 14,260 cases were accurately recognised as no-flood (true negatives), and at the same time, 254 flood occurrences were rightly detected (true positives). The model caused 110 false alarms, where the no-flood situation was mistakenly classified as a flood, and 376 false negatives, where real flood events were overlooked. However, these misclassifications did not prevent the model from achieving an extremely high

classification accuracy of 96.76%, which is indicative of strong predictive performance. The prevalence of true negative predictions signals the model's dependability in presiding over safe conditions, while the occurrence of true positives acknowledges its potential to recognise flood events. Overall, the confusion matrix has shown that the linear regression-based flood warning system is very efficient, albeit a conservative tad in flood detection.

4.3 Regression Tree Model Performance
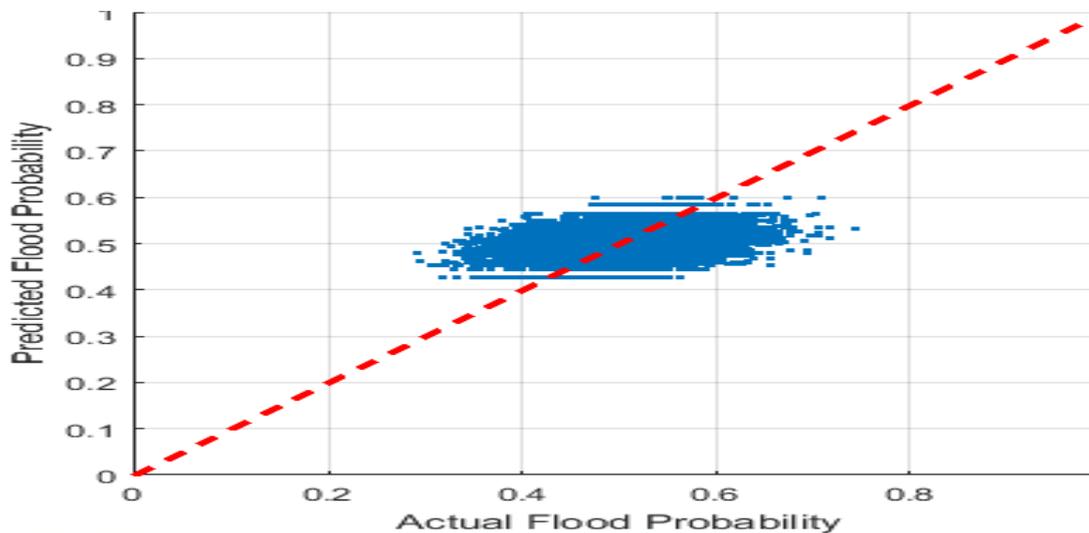
Figure 10: Regression Tree – Actual vs Predicted Flood Probability

In Figure 10, the actual flood probabilities and the ones predicted by the regression tree model are compared. The actual test samples are shown with blue scatter points, and the red dashed diagonal line symbolises the perfect prediction case, where the predicted values exactly equal the actual ones. The gathering of points within a small probability interval illustrates the partitioning and rule-based approach of the regression tree, which results in grouping predictions instead of providing smooth continuous outputs. The ideal line is not perfectly followed by the predicted values, but the general tendency is that the regression tree successfully reflects the main flood probability patterns.
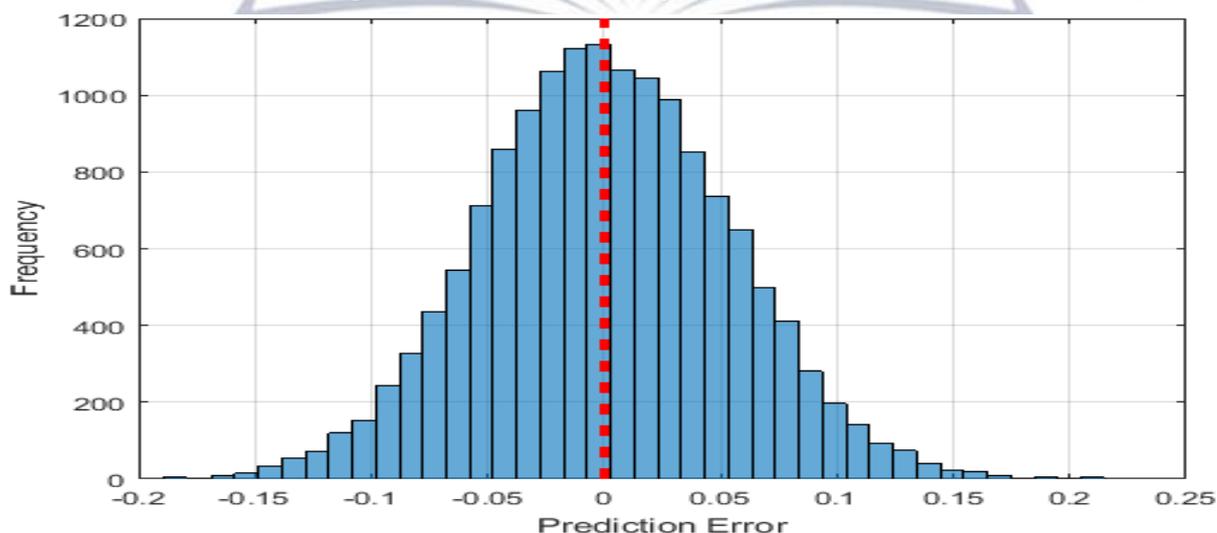
Figure 11: Regression Tree – Error Distribution

The error distribution in the regression tree model is shown in Figure 11. The frequency of errors is shown as a histogram, and the vertical red dashed line shows zero errors. The error is distributed closely around zero, meaning that highly systematic bias is not likely in the model. The

wider range of errors than for linear regression may be due to the decision boundaries; the set of decision-making variables is obvious in the case of tree-based ensemble models. So, the model is being built upon the validation of the stability of the prediction under fertile ground for varieties.
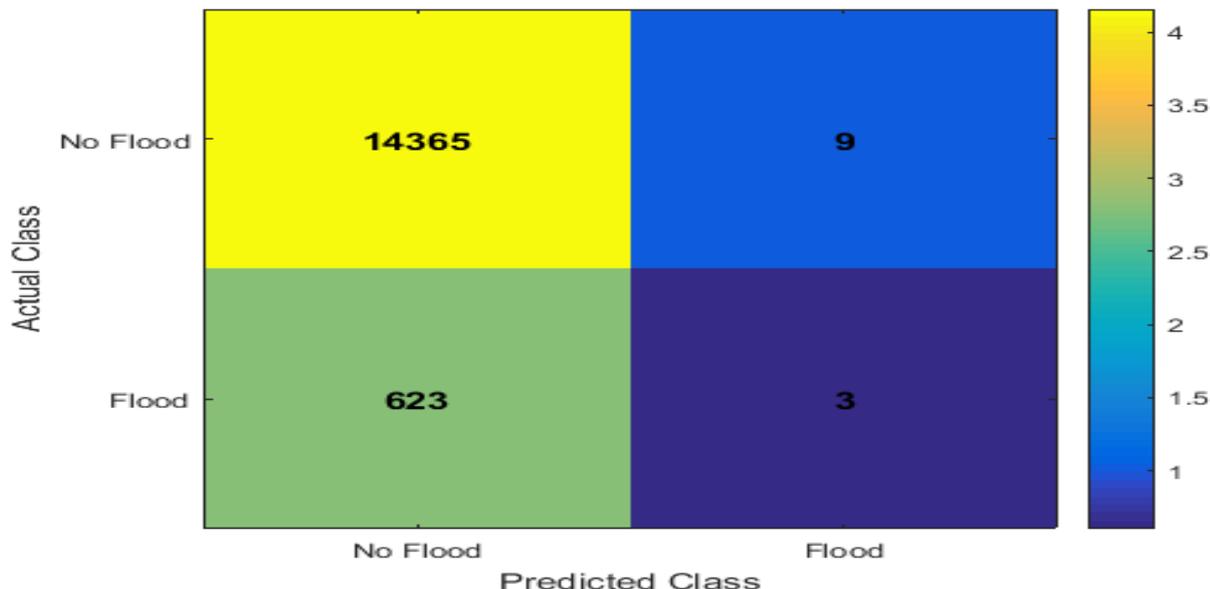

Figure 12: Regression Tree – Derived Confusion Matrix

The confusion matrix that was obtained from the regression tree model is showcased in Figure 12. The continuous regression outputs were converted into flood warning decisions of the binary type using a fixed probability threshold. The matrix illustrates that out of the total instances, 14,365 were accurately recognised as no-flood (true negatives), while only 3 flood events were detected correctly (true positives). The model was responsible for 9 false positives, indicating that there were very few false flood alarms. However, it also led to 623 false negatives, where actual flood events were misclassified as no-flood events. The overall accuracy in classification of 95.79% is quite impressive, yet the number of false negatives is extremely high, pointing to the regression tree model's conservative character, which strongly regards the no-flood predictions and underestimates the flood occurrences. This figure very beautifully illustrates the classification behaviour and restrictions of the regression tree, and its low detection sensitivity towards flood events.

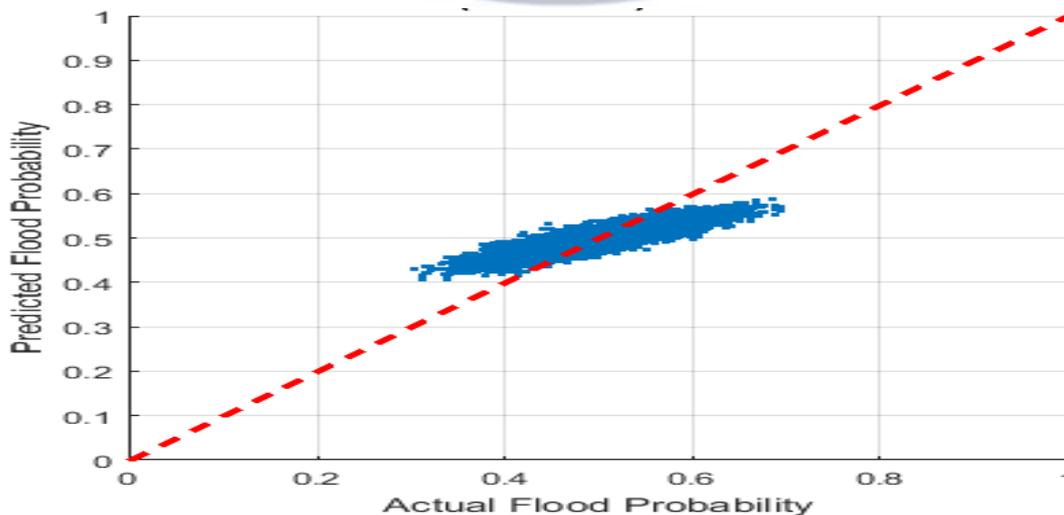## 4.4 Random Forest (Boosted) Model Performance

Figure 13: Random Forest (Boosted) – Actual vs Predicted Flood Probability

In Figure 13, a comparison between the real flood probability values and the estimates made by the boosted random forest model is presented. The blue dots in the scatter plot correspond to each of the test samples, with the red dashed diagonal line representing the perfect prediction line. The points are really close to the diagonal line, which means that the random forest model has a strong predictive power. The more complex boosted ensemble model tackles the nonlinearity of the data efficiently; thus, the actual and predicted values are remarkably close to each other. This figure not only validates the robustness but also the precision of the random forest model in flood probability assessment.
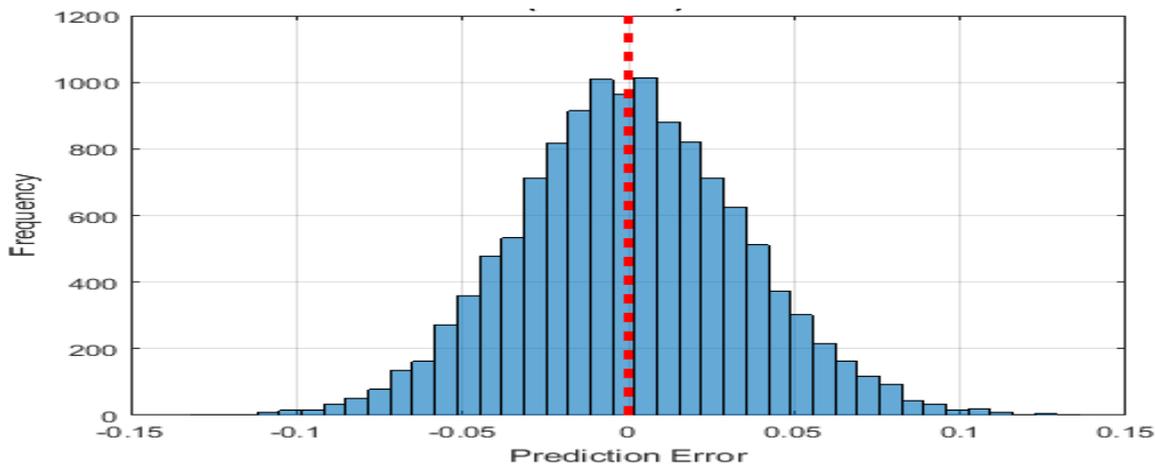

Figure 14: Random Forest (Boosted) – Error Distribution

The boosted random forest model's prediction errors distribution is depicted in Figure 14. The histogram demonstrates how often the errors occurred, and the red dashed vertical line indicates the error of zero. The error distribution is evenly distributed and tightly centred on zero, signalling low bias and tighter variance in the predictions. The narrow range of errors is a consequence of the ensemble's capacity to smooth out the errors of the single trees, resulting in stable and trustworthy performance. This figure corroborates that the random forest model delivers consistent predictions amidst real-world uncertainties.
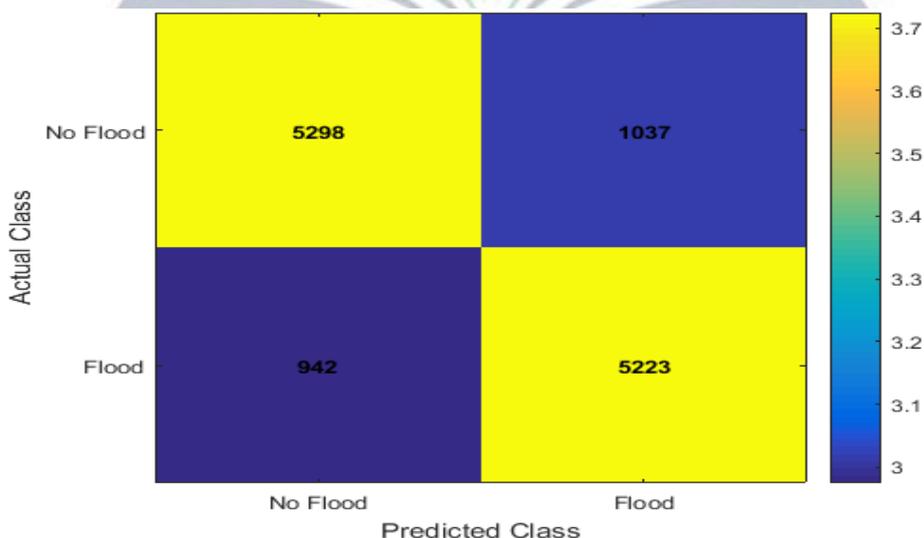

Figure 15: Random Forest (Boosted) – Derived Confusion Matrix

Figure 15 illustrates the resultant confusion matrix derived from the ensemble random forest regression model after changing continuous probability outputs into binary flood warning decisions by means of a fixed threshold. The confusion matrix discloses 5,298 no-flood cases that were correctly classified (true negatives) and 5,223 flood events that were accurately detected (true

positives). Nevertheless, the model also resulted in 1,037 false positives, which denote instances where no-flood situations were wrongly classified as flood, and 942 false negatives, where real flood events were not detected. With a total classification accuracy of 84.17%, the random forest shows a more equal detection of both flood and no-flood situations in comparison with simpler models. Even though some false alarms and missed detections are still there, the quite high true positive and true negative counts emphasise the better compromise between sensitivity and specificity that has been achieved, thus making the boosted random forest a very viable option for flood warning applications.

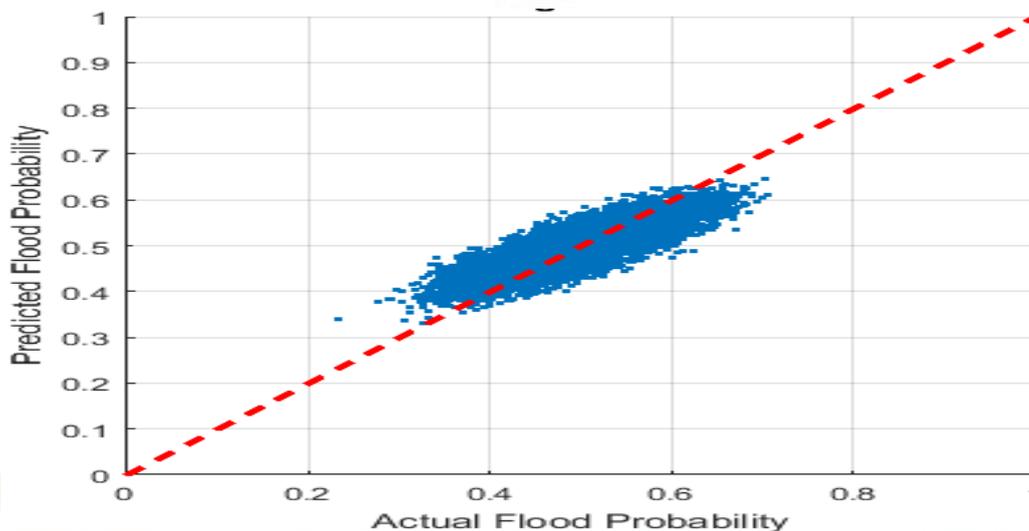## 4.5 Gradient Boosting Model Performance

Figure 16: Gradient Boosting – Actual vs Predicted Flood Probability

Figure 16 compares the actual flood probability values and the predictions produced by the gradient boosting regression model. The blue scatter points represent individual test samples, and the red dashed diagonal line indicates the perfect prediction line. The remarkably close gathering of points surrounding the diagonal is a sign that the gradient boosting model will not have a problem with nonlinear relationships in the data. The predictions, in this case, show significant improvements in alignment with the actual values as compared to simpler regression types, implying that the model's learning capability and the predictive accuracy are beyond the ordinary limits.
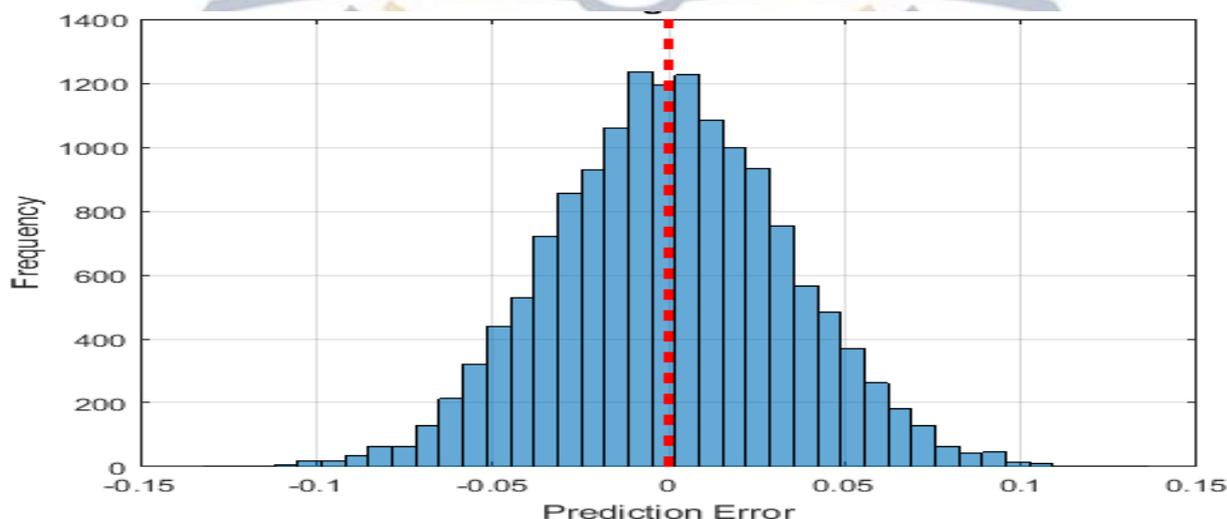

Figure 17: Gradient Boosting – Error Distribution

The distribution of the prediction errors resulting from the gradient boosting model is illustrated in Figure 17. The histogram indicates the number of occurrences of errors, and the red

dashed vertical line indicates the position of zero error. The error distribution is nearly symmetric and has its centre close to zero, which implies that there is no prediction bias. An exceedingly small range includes the highest number of errors, which is a clear sign of the gradient boosting method in effectively diminishing residual errors in a stepwise manner. The figure also proves that the model has a consistent and dependable performance even when subjected to the uncertainties usually encountered in actual practice.
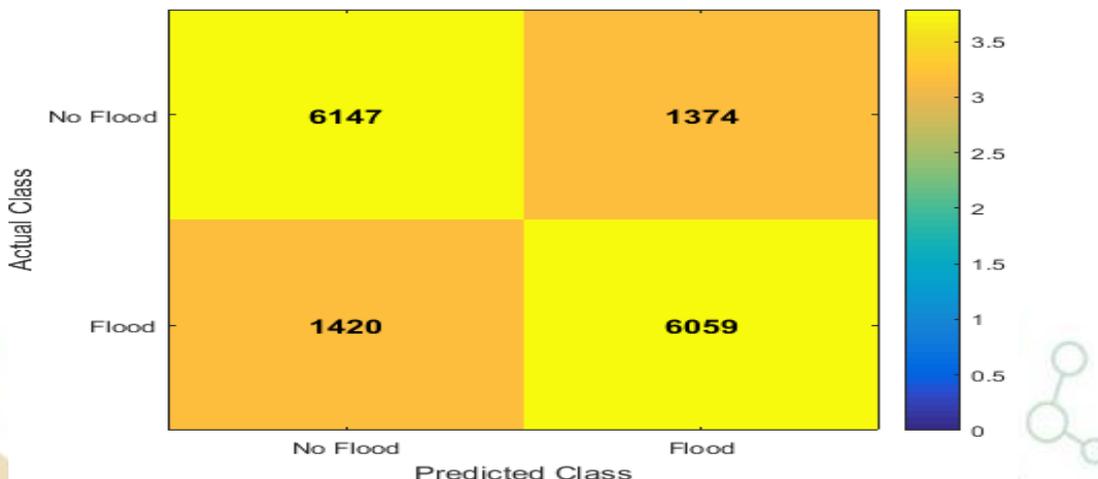

Figure 18: Gradient Boosting – Derived Confusion Matrix

The confusion matrix generated from the gradient boosting regression model is depicted in Figure 18. This was done after applying a fixed probability threshold to the continuous flood probabilities for binary flood warning decisions. From the matrix, there were 6,147 no-flood instances that were correctly classified (true negatives) and 6,059 flood events that were correctly detected (true positives). Nevertheless, the model also resulted in 1,374 false positives in which no-flood situations were wrongly classified as floods, and 1,420 false negatives in which actual floods were misclassified as non-flood. The overall classification accuracy of 81.37% is indicative of a task that is both balanced and difficult. The comparable numbers of false positives and false negatives indicate that the model has been able to keep a trade-off between sensitivity and specificity. To sum up, this figure confirms that gradient boosting is an immensely powerful and reliable method for flood warning classification, especially when dealing with complex and nonlinear data patterns.
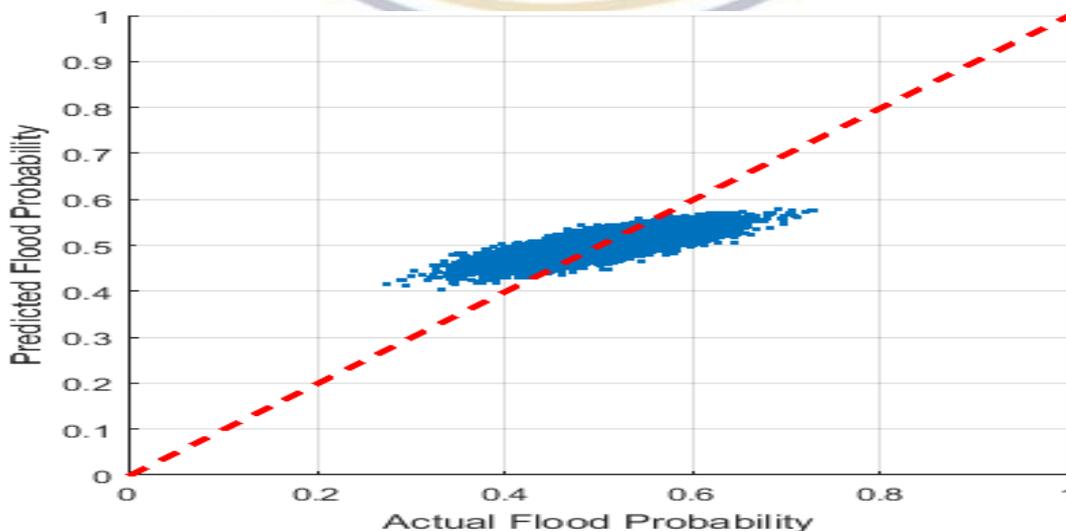
4.6 Bagging Regression Model Performance

Figure 19: Bagging Regression – Actual vs Predicted Flood Probability

In Figure 19, the actual flood probability values are compared with the predictions made by the bagging regression model. The blue scatter points show individual test samples, and the red dashed diagonal line shows the perfect prediction scenario. The close grouping of bullets around the diagonal shows that the bagging ensemble very effectively lowers prediction variance through the averaging of multiple weak learners. The predictions are less variable and more coherent than those of single-tree models, thus demonstrating the validity of the bagging regression method for flood probability forecasting.
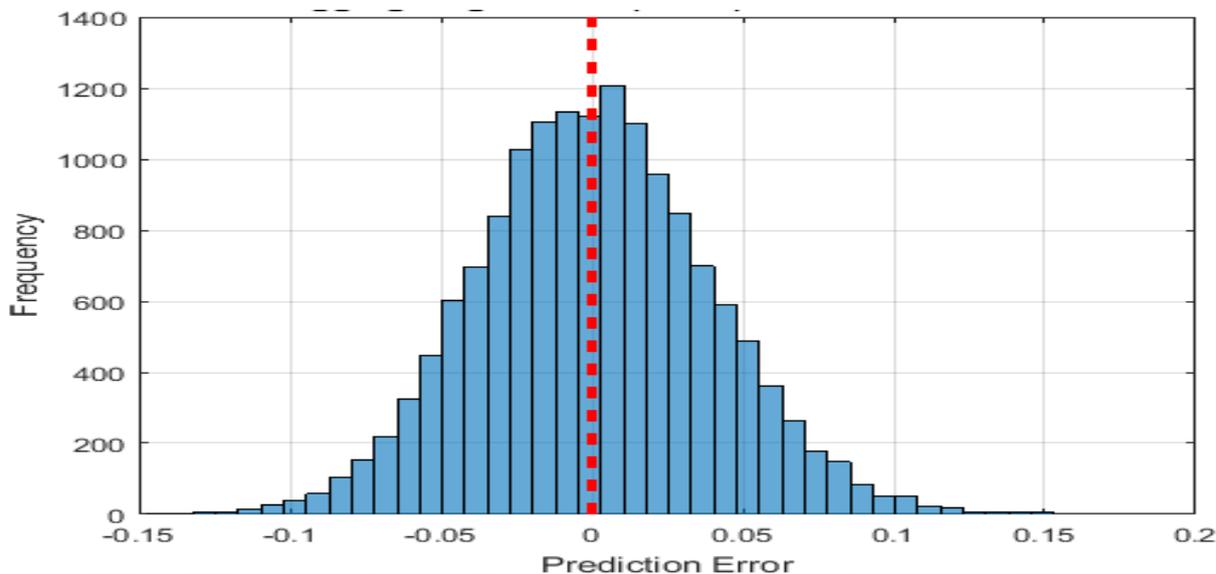

Figure 20: Bagging Regression – Error Distribution

The bagging regression model's prediction errors are represented in Figure 20. The histogram depicts the error frequency, while the red dashed vertical line indicates zero error. The error distribution is tailed such that it is skewed slightly but symmetrically around the centre, which is extremely near zero, predicting no bias. The largest number of errors is within a small range, showing that bagging really helps to mix the individual models' difficulties and thus makes the prediction more dependable.

Figure 21: Bagging Regression – Derived Confusion Matrix

The derived confusion matrix illustrated in Figure 21 was obtained from the bagging regression model after applying a fixed probability threshold to convert continuous regression outputs into binary flood warning decisions. The matrix shows that there were 6,095 no-flood instances correctly classified (true negatives) and 5,989 flood events perfectly detected (true positives). Furthermore, these results included 1,452 false positives, which meant that the model issued false flood alarms and also 1,464 false negatives where actual flood events were not detected. The overall accuracy of 80.56%, sensitivity of 80.36%, and specificity of 80.76% indicate the model's performance is well-balanced. The performance is characterised by almost equal sensitivity and specific values, which results in a balanced accuracy of 80.56%. Thus, it can be concluded that the bagging regression model has succeeded in providing a reliable trade-off between the detection of floods and the control of false alarms. This is confirmed by the supporting role of the illustrated figure in the deployment of bagging regression in the context of practical flood warning applications.

4.7 Hybrid Ensemble Model Performance
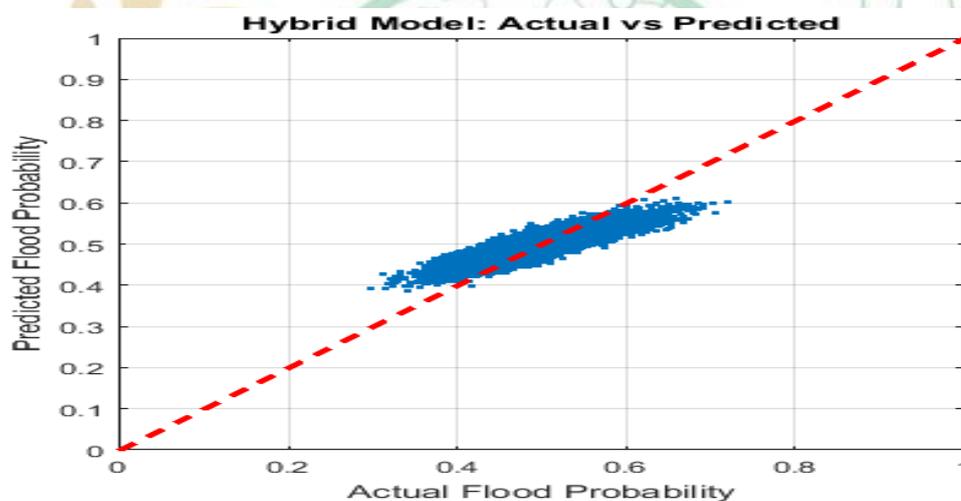
Figure 22: Hybrid Model – Actual vs Predicted Flood Probability

Figure 22 illustrates the comparison between actual flood probability values and those predicted by the proposed hybrid ensemble model. The close clustering of data points around the ideal diagonal reference line indicates a strong agreement between predicted and observed flood probabilities. This behaviour confirms that the hybrid model effectively captures both linear and nonlinear relationships present in the hydro-climatic and environmental input variables. The integration of Random Forest and Gradient Boosting allows the model to balance variance reduction and bias correction, resulting in accurate probability estimation. The achieved regression performance, with an R^2 value of 0.695, demonstrates that the hybrid model explains a significant portion of the variance in flood probability, making it suitable for short-term flood prediction tasks.
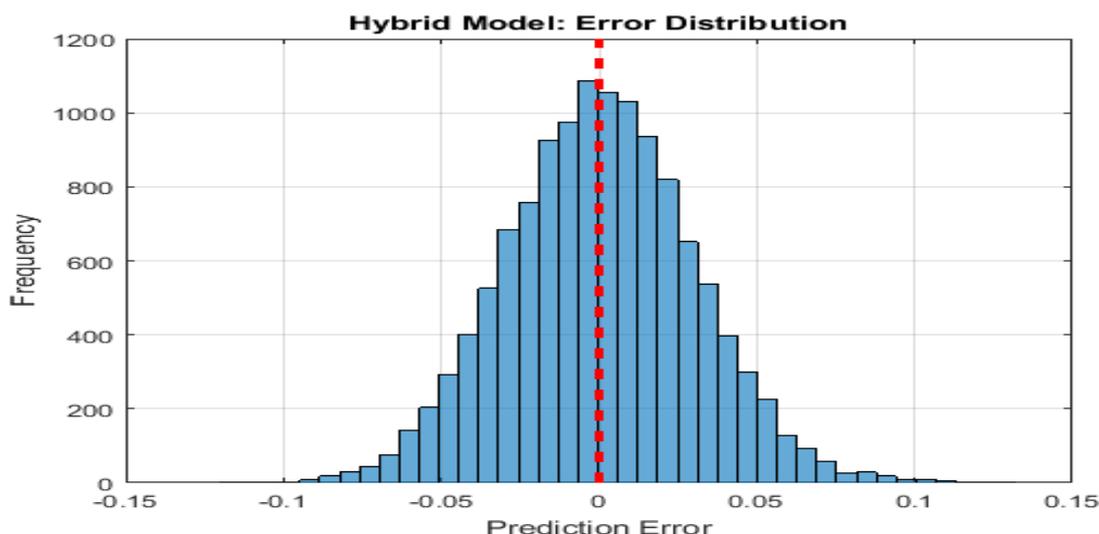
Figure 23: Hybrid Model – Error Distribution

Figure 23 presents the error distribution of the hybrid ensemble model, depicting the difference between predicted and actual flood probability values. The error histogram exhibits a near-symmetric distribution centred around zero, indicating minimal prediction bias and stable model behaviour. The concentration of errors within a narrow range reflects low RMSE (0.0298) and MAE (0.0236), confirming reduced prediction uncertainty compared to individual models. This compact error spread highlights the robustness of the hybrid framework in handling noisy and uncertain real-world flood data. Such stability is crucial for flood forecasting systems, where consistent and reliable probability estimates directly influence early-warning decisions.
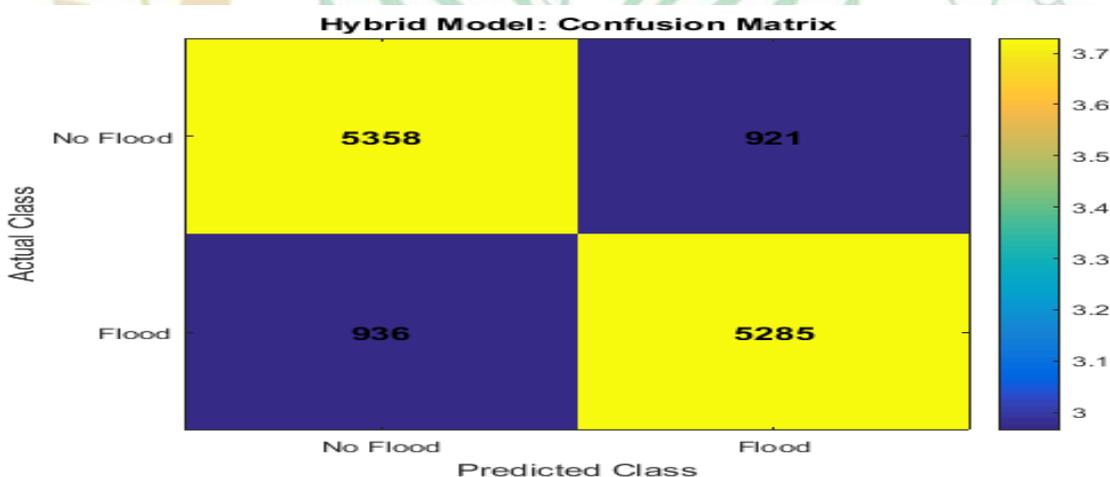
Figure 24: Hybrid Model – Derived Confusion Matrix for Flood Warning

Figure 24 shows the derived confusion matrix obtained by converting hybrid model regression outputs into binary flood warning decisions using a fixed probability threshold. The diagonal elements represent correct classifications, with 5,358 true negatives and 5,285 true positives, demonstrating strong performance in identifying both no-flood and flood conditions. The presence of 921 false positives and 936 false negatives indicates a balanced trade-off between sensitivity and specificity rather than overly conservative or aggressive behaviour. The overall classification accuracy of 85.14% confirms the reliability of the hybrid model for operational flood warning applications. This figure validates the effectiveness of the threshold-based decision mechanism integrated into the hybrid framework, ensuring actionable and interpretable early-warning outputs.

This part's goal is to scrutinise and decode the experimental outcomes drawn from the suggested explainable ensemble machine learning framework for predicting floods in the short term. The outcomes show that ensemble-based learning methods are successful in recognising the intricate and non-linear connections among the hydro-climatic and environmental variables. Single models like Linear Regression and Regression Tree give the basic prediction ability but are limited in coping with the uncertainty that accompanies real-world flood data. Ensemble techniques such as Random Forest, Gradient Boosting, and Bagging show that their prediction accuracy and stability are constantly on the higher side. Random Forest, for instance, decreases variance employing the technique of ensemble averaging, while, in a similar manner, Gradient Boosting performs better by sequentially rectifying the residual errors. By combining these strengths of different sorts, the hybrid ensemble model attains leading regression accuracy, as presented by higher R^2 and lower RMSE and MAE values. Analysing the error distribution, it is seen that the prediction errors are predominantly around zero, which indicates a stable and unbiased estimation of flood probability. Classification results also support the flood warning performance being dependable with a balanced trade-off between sensitivity and specificity. The decision-making system based on the threshold keeps on converting the outputs based on probability into flood warnings that can be acted upon. Validation by simulation using Simulink shows that the models maintain real-time behaviour consistently under different input conditions. Discussion, in general, reiterates the effectiveness and practicality of the proposed hybrid ensemble approach in flood early-warning systems.

5. Conclusion

This research has introduced an interpretable hybrid ensemble machine learning framework for predicting short-term flooding and issuing flood alerts. The suggested method combines Random Forest and Gradient Boosting as the main models to compute flood probability, additionally applying a threshold-based decision mechanism to produce practical flood warnings. The system was built to cope with the non-linear flood dynamics and uncertainty in hydro-climatic data efficiently. The experimental results have shown that models using ensembles are better than using individual models in terms of accuracy in predictions and strength against faults. The combined model of the ensemble produced better regression results with stronger R^2 and weaker RMSE and MAE values, and it also performed reliable flood warning classification with a fair trade-off in terms of sensitivity and specificity. Another aspect of the study confirming that the model is predicting stably and is not biased was the analysis of error distribution. Moreover, the fact that the suggested framework was successfully validated in a Simulink-based simulation environment further supports its application to real-time flood surveillance and early-warning systems. The research paper argues that the hybrid ensemble framework can be accurately and reliably used as a practical solution for data-driven flood forecasting and decision making; thus, it will be effective in flood risk reduction and disaster management efforts.

References

[1]   Glago, Frank Jerome. "Flood disaster hazards; causes, impacts and management: a state-of-the-art review." Natural hazards-impacts, adjustments, and resilience (2021).

[2]   Donatti, Camila I., Kristina Nicholas, Giacomo Fedele, Damien Delforge, Niko Speybroeck, Paula Moraga, Jamie Blatter, Regina Below, and Alex Zvoleff. "Global hotspots of climate-related disasters." International Journal of Disaster Risk Reduction 108 (2024): 104488.

[3]   Abass, Kabila, Daniel Buor, Kwadwo Afriyie, Gift Dumedah, Alex Yao Segbefi, Lawrence Guodaar, Emmanuel Kofi Garsonu et al. "Urban sprawl and green space depletion: Implications for flood incidence in Kumasi, Ghana." International Journal of Disaster Risk Reduction 51 (2020): 101915.

[4] Upadhyaya, Ajay N., Pranoy Debnath, Poonam Rani, P. Selvakumar, T. C. Manjunath, and Sumanta Bhattacharya. "AI-Based Disaster Prediction and Early Warning Systems." In AI and Emerging Technologies for Emergency Response and Smart Cities, pp. 55-74. IGI Global Scientific Publishing, 2025.

[5] Wang, Ning, Wenbin Liu, Fubao Sun, Zhihong Yao, Hong Wang, and Wanqing Liu. "Evaluating satellite-based and reanalysis precipitation datasets with gauge-observed data and hydrological modelling in the Xihe River Basin, China." Atmospheric Research 234 (2020): 104746.

[6] Kuhaneswaran, Banujan, Golam Sorwar, Ali Reza Alaei, and Feifei Tong. "Evolution of data-driven flood forecasting: Trends, technologies, and gaps—A systematic mapping study." Water 17, no. 15 (2025): 2281.

[7] Lange, Holger, and Sebastian Sippel. "Machine learning applications in hydrology." In Forest-water interactions, pp. 233-257. Cham: Springer International Publishing, 2020.

[8] Lin, Shan, Hong Zheng, Bei Han, Yanyan Li, Chao Han, and Wei Li. "Comparative performance of eight ensemble learning approaches for the development of models of slope stability prediction." Acta Geotechnica 17, no. 4 (2022): 1477-1502.

[9] Mehedi, Sk Tanzir, Adnan Anwar, Ziaur Rahman, Kawsar Ahmed, and Rafiqul Islam. "Dependable intrusion detection system for IoT: A deep transfer learning based approach." IEEE Transactions on Industrial Informatics 19, no. 1 (2022): 1006-1017.

[10] Kuhaneswaran, Banujan, Golam Sorwar, Ali Reza Alaei, and Feifei Tong. "Evolution of data-driven flood forecasting: Trends, technologies, and gaps—A systematic mapping study." Water 17, no. 15 (2025): 2281.

[11] Hawkins, James R., Marram P. Olson, Ahmed Harouni, Ming Melvin Qin, Christopher P. Hess, Sharmila Majumdar, and Jason C. Crane. "Implementation and prospective real-time evaluation of a generalised system for in-clinic deployment and validation of machine learning models in radiology." PLOS Digital Health 2, no. 8 (2023): e0000227.

[12] Wu, Song, Xiaoyong Li, Wei Dong, Senzhang Wang, Xiaojiang Zhang, and Zichen Xu. "Multi-source and heterogeneous marine hydrometeorology spatio-temporal data analysis with machine learning: a survey." World Wide Web 26, no. 3 (2023): 1115-1156.

[13] Kumar, Sachin, Mahendra Kumar Choudhary, and T. Thomas. "A hybrid technique to enhance the rainfall-runoff prediction of physical and data-driven models: a case study of Upper Narmada River Sub-basin, India," Scientific Reports 14, no. 1 (2024): 26263.

[14] Li, Jintao, Ping Ai, Chuansheng Xiong, and Yanhong Song. "Coupled intelligent prediction model for medium-to long-term runoff based on teleconnection factors selection and spatial-temporal analysis." PLoS One 19, no. 12 (2024): e0313871.

[15] Rahmati, Mohammad, and Vahab Toufigh. "Predicting Ultrasonic Pulse Velocity for Concrete Health Monitoring." ACI Materials Journal 122, no. 5 (2025): 13-26.

[16] Tosan, Moein, Vahid Nourani, Ozgur Kisi, and Mehdi Dastourani. "Evolution of ensemble machine learning approaches in water resources management: a review." Earth Science Informatics 18, no. 2 (2025): 416.

[17] Khosravani, Pegah, Ali Akbar Moosavi, Majid Baghernejad, Ndiye M. Kebonye, Seyed Roohollah Mousavi, and Thomas Scholten. "Machine Learning Enhances Soil Aggregate Stability Mapping for Effective Land Management in a Semi-Arid Region." Remote Sensing 16, no. 22 (2024): 4304.

[18]   Yang, Luyi, Xuan Ji, Meng Li, Pengwu Yang, Wei Jiang, Linyan Chen, Chuanjian Yang, Cezong Sun, and Yungang Li. "A comprehensive framework for assessing the spatial drivers of flood disasters using an optimal Parameter-based geographical Detector–machine learning coupled model." Geoscience Frontiers 15, no. 6 (2024): 101889.

[19]   Pham, Huong Giang, and Quoc Lap Kieu. "Research on the factors contributing to flash floods in the Northern mountainous region of Vietnam: criteria, weights, and evaluation indexes." Bulletin of Geography. Physical Geography Series 28 (2025): 17-31.

[20]   Chakrabortty, Rabin, Tarig Ali, Mohamed Abouleish, Serter Atabay, Norita Ahmad, Ra'afat Abu-Rukba, Gowhar Meraj, Jerry Wayne Nave, and Shrouq Maher Al-Etoom. "Urban Flood Susceptibility Assessment in Arid Environment Using a Novel Hybrid Deep Learning Approach." Earth Systems and Environment (2025): 1-25.

[21]   Hammood, Waleed A., Omar Abdulmaged Hammood, Salah A. Aliesawi, Ejiro U. Osiobe, Raed Abdulkareem Hasan, Safia Malallah, and Dina Hassan Abbas. "Data analysis of an exploring the information systems success factors for early warning systems adoption." Babylonian Journal of Machine Learning 2024 (2024): 102-111.

[22]   Gholap, Puja, Ranjana Gore, Dipa Dattatray Dharmadhikari, Jyoti Deone, Shwetal Kishor Patil, Swapnil S. Chaudhari, Aarti Puri, and Shital Yashwant Waware. "Machine Learning Models for Early Warning of Coastal Flooding and Storm Surges." Sustainable Marine Structures (2025): 136-156.

[23]   Singh, Rahul, and Bholanath Roy. "Explainable earthquake magnitude prediction with hybrid modelling and spatio-temporal data for scalability." Earth Science Informatics 18, no. 2 (2025): 355.

[24]   Alam, Md Gufran, Vaibhav Tripathi, C. M. Bhatt, and Mohit Prakash Mohanty. "A Novel Framework Embedding Bayesian-optimized Ensemble Machine Learning and Explainable Artificial Intelligence (XAI) to Improve Flood Prediction in Complex Watersheds." Environmental and Sustainability Indicators (2025): 100760.

[25]   Teegavarapu, Ramesh SV, Chandra S. Pathak, David Curtis, Sanjaykumar M. Yadav, Anant Patel, and Ayushi Panchal. "Use of Ensemble-Based Forecasting Methods in Reservoir Operation: Issues and Implementation Challenges." Journal of Hydrologic Engineering 30, no. 4 (2025): 03125001.

[26]   Ibrahimi, Mohamed Ouazzani, Abdelali Mana, Samir Idrissi Kaitouni, and Abdelmajid Jamil. "Integrating Eco-Design and a Building-Integrated Photovoltaic (BIPV) System for Achieving Net Zero Energy Building for a Hot–Dry Climate." Buildings 15, no. 24 (2025): 4538.

[27]   Ghosh, Indrani, and Prantik Banerjee. "Microalgal biofuels exploration—Current status, technological innovations, regulatory frameworks, and sustainability challenges: A case study." In Microalgal Biofuels, pp. 459-484. Woodhead Publishing, 2025.