Forecasting floods with machine learning methods

ZOI MOUSTAKI, EVANGELIA N. PETRAKI
Department of Economics
National and Kapodistrian University of Athens
Sofokleous 1, 10559, Athens
GREECE

Abstract: - The current paper investigates the possibility of flood risk prediction with machine learning methods. Global warming due to climate change contributes to the increase in the frequency of flooding events, which are a significant consequence of the climate crisis. Floods are caused by a combination of natural factors (such as heavy rainfall and geomorphology) and human activities, with urbanization being the most significant among them. Their impacts extend to many sectors, such as the economy, the environment, infrastructure, and public health. The current study aims to develop a reliable flood risk prediction system using machine learning techniques. The analysis is based on meteorological, hydrological, geographical, socio-economic and historical flood data from regions of India. Early flood forecasting can contribute substantially to taking preventive measures and mitigating their impacts.

Key-Words: - Flood risk prediction, Climate change, Machine learning, India, Urbanization, Preventive Measures

Received: March 5, 2026. Revised: April 3, 2026. Accepted: May 7, 2026. Published: July 13, 2026.

1 Introduction

Climate change is a global phenomenon that is increasingly making its presence felt in people's daily lives and is expected to become even more intense and frequent. However, over time, actions have been taken to address it. Initially, in 1988, the establishment of the Intergovernmental Panel on Climate Change (IPCC) by the World Meteorological Organization (WMO) and the United Nations Environment Programme (UNEP), marked the global recognition of the problem and the formulation of policy action [1] In 1997, the Kyoto Protocol was adopted, aiming to reduce greenhouse gas emissions by industrialized countries [2]. Then, in 2015, the Paris Agreement was signed to limit global warming to below 1.5 ºC compared to pre-industrial levels [3]. Finally, by 2030, global emissions are expected to be reduced by 43%, according to the projections and targets of the United Nations Framework Convention on Climate Change (UNFCCC) and the Intergovernmental Panel on Climate Change (IPCC), with the aim of limiting temperature rise to below 1.5 ºC [4].

Although the impacts of climate change are felt in various sectors, such as health and agriculture, the scientific community has paid particular attention to environmental consequences, with an emphasis on the increasing frequency and intensity of flooding events [5]. This is because global warming affects the water cycle, increasing evaporation, intensifying rainfall, and contributing to river and dam overflows [6], [7]. Floods are classified, based on their origin and the weather and/or geographic conditions that cause them, into river floods, flash floods, coastal floods, storm surges and inland floods [8].

The causes of floods arise from both natural and anthropogenic factors. Natural parameters include intense and prolonged rainfall, the geomorphology of the area (altitude, land slope, proximity to rivers or seas), snowfall and rapid ice melting, as well as seismic, volcanic and geological phenomena related to erosion and instability [9], [10], [11], [12], [13], [14], [15]. Similarly, anthropogenic factors consist of extensive urbanization, uncontrolled construction, inadequate maintenance of dams and hydraulic infrastructure, as well as deforestation, which reduces the soil's capacity to absorb and retain water [16], [17], [18].

Floods have severe impacts on various sectors. Economically, they cause extensive damage to critical infrastructure and equipment, increasing the financial burden of recovery for those affected [19]. Environmentally, they contribute to soil subsidence, sediment transport, and the destruction of vegetation [20]. In terms of health, they pose risks such as drowning, injuries, and the spread of diseases [21], while socially, they negatively affect mental health due to the loss of shelter, with subsequent financial insecurity [22]. Finally, they cause significant damage to transportation systems and other critical infrastructure [23].

Therefore, floods cause serious impacts across many aspects of life and the implementation of effective prevention measures is essential. Consequently, the construction of flood protection infrastructure is imperative, both in areas that have already been affected and in regions vulnerable due to climate change. Key projects and interventions that can be implemented include the construction and upgrading of dams, storage tanks and levees [24], as well as the enhancement of forecasting and early warning mechanisms through technologies, such as the aquaFLOOD platform, local warning systems in India, the Soil and Water Assessment Tool (SWAT) model, etc. [25], [26], [27]. Furthermore, constructing green roofs with plants and soil, along with using permeable pavements, helps in retaining rainfall, promotes its infiltration into the soil and reduces surface runoff [28]. Regular cleaning of drains and sewage systems is also a critical preventive measure [29], as is the revision of urban planning to avoid construction in high-risk areas [30]. Finally, sustainable flood management requires international cooperation to reduce carbon dioxide emissions and address climate change.

In addition to preventive measures and infrastructure interventions, the forecasting of flood events through advanced computational models constitutes a critical tool in flood management. The development and implementation of such predictive models allows for the early identification of risks and helps authorities to design and implement strategies that mitigate or eliminate the impacts. The accuracy of forecasts is enhanced by analyzing the parameters that affect the occurrence of floods, combined with modern machine learning techniques [31]. However, despite technological progress, it must be taken into account that unpredictable factors may affect the accuracy of predictions.

Thus, the present study aims to evaluate and compare the effectiveness of various machine learning algorithms for predicting flood events in different regions of India. In contrast to previous studies that focus exclusively on hydrological and meteorological features, this study incorporates socio-economic and infrastructural factors, offering a more comprehensive approach. This methodology contributes to a better understanding of the model results and supports informed decision-making for effective flood risk management. By using data from these specific regions, the current research aims to identify the most accurate and reliable models to enhance early warning systems and improve overall flood risk management.

2 Description of the dataset

The dataset used for developing flood prediction models across a wide range of regions in India includes meteorological, hydrological, geographical, socio-economic, and historical flood data. These help in understanding the causes of floods and in creating more reliable predictions. The dataset, consisting of 14 columns and 10,000 rows, is available on Kaggle from Jaya Surya (https://www.kaggle.com/datasets/s3programmer/flood-risk-in-india).

Specifically, the dataset contains the following variables:

- **Latitude** (float): Represents the geographic latitude of the observation site.
- **Longitude** (float): Indicates the geographic longitude of the observation location.
- **Rainfall_mm** (float): Represents the volume of rainfall (in millimeters) recorded at a particular location.
- **Temperature_C** (float): Indicates the temperature in degrees Celsius (°C) at the corresponding location.
- **Humidity_pct** (float): Shows the percentage of relative humidity at the observed area.
- **River_Discharge_m3_s** (float): Denotes the volume of water flowing per second (in cubic meters).
- **Water_Level_m** (float): Indicates the height of the water surface (in meters) at a specified location.
- **Elevation_m** (float): Represents the height of the location above mean sea level (in meters).
- **Land_Cover** (categorical): Describes the type of land cover in the area. Categories include Urban, Forest, Agricultural, Water Body, and Desert.
- **Soil_Type** (categorical): Describes the soil composition. Categories include Sandy, Clay, Silt, Loam, and Peat.
- **Population_Density** (float): Indicates the population density, measured in individuals per square kilometer.
- **Infrastructure** (binary: 0/1): Specifies the presence (1) or absence (0) of flood control infrastructure.
- **Historical_Floods** (binary: 0/1): Indicates whether the area has experienced flood events in the past (1: yes, 0: no).

- **Flood_Occurred** (binary: 0/1): The target variable, denoting the occurrence (1) or non-occurrence (0) of a flood event in the given area.

The following table (Table 1) presents the summary descriptive statistics, including count, mean, standard deviation, minimum, and maximum values for each numerical feature:

	Latitude	Longitude	Rainfall (mm)	Temperature (°C)	Humidity (%)	River Discharge (m³/s)
count	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000
mean	22.330627	82.631366	150.015118	29.961401	59.749104	2515.722946
std	8.341274	8.389542	86.032127	8.669838	23.142734	1441.706442
min	8.000337	68.004575	0.014437	15.000166	20.001339	0.042161
25%	15.143537	75.364428	76.124373	22.405717	39.541778	1284.782376
50%	22.283330	82.671007	150.620428	30.000907	59.497375	2530.451944
75%	29.460184	89.937897	223.402156	37.413488	80.038163	3767.229862
max	36.991813	96.997820	299.970293	44.993681	99.997772	4999.698480

Water Level (m)	Elevation (m)	Population Density	Infrastructure	Historical Floods	Flood Occurred
10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000
5.017881	4417.138177	5021.468442	0.502000	0.498700	0.505700
2.876579	2530.245421	2882.591520	0.500021	0.500023	0.499993
0.002701	1.150340	2.289000	0.000000	0.000000	0.000000
2.538847	2229.681903	2491.766601	0.000000	0.000000	0.000000
5.042094	4417.199761	5074.392879	1.000000	0.000000	1.000000
7.524692	6616.729066	7474.228752	1.000000	1.000000	1.000000
9.996899	8846.894877	9999.169530	1.000000	1.000000	1.000000

Table 1. Summary Descriptive Statistics of Numerical Features.

Summarizing the numerical characteristics related to flood risk, the following observations are made:

- The average rainfall is 150 millimeters, with a range of 0 to 300 millimeters, demonstrating the variability that influences flood risk.
- Temperature varies between 15 °C and 45 °C, with a mean of 30 °C, promoting evaporation and affecting water management.
- Humidity ranges from 20% to 100%, with an average of 60%, and higher humidity levels are associated with an elevated risk of flooding.
- The average water level is 5 meters, reaching up to 10 meters, significantly affecting low-lying areas.
- The mean population density is 5,021 inhabitants per square kilometer, ranging from 2 to 10,000, with densely populated areas exhibiting higher human activity that can intensify flood risk.
- Historical flood occurrences have a mean of 0.5, indicating that approximately 50% of the regions have experienced past flooding events, highlighting potential vulnerability in the future.

Summary descriptive statistics of the categorical features are provided in Table 2:

	Land Cover	Soil Type
count	10000	10000
unique	5	5
top	Water Body	Peat
freq	2046	2052

Table 2. Summary Descriptive Statistics of Categorical Features.

The variables “Land Cover” and “Soil Type” each contain 10,000 records and consist of 5 distinct categories. The highest frequency in “Land Cover” corresponds to Water Body (2,046 records) and in “Soil Type” to Peat (2,052 records), indicating that the study area is largely covered by water bodies and wetlands.

3 Flood Prediction Using Machine Learning Methods

For each method, we used two ways to split the dataset: the Train-Test Split with test_size=0.2 and test_size=0.3, and K-Fold Cross-Validation. Then, we analyze which splitting approach performs best for each method.

The machine learning methods we use are the following:

3.1 Logistic Regression

K-Fold Cross-Validation

In this evaluation method, we divide the dataset into k equal parts, where k is the number of folds that achieves a balance between the accuracy of the estimate and the computational cost. In this analysis, it was found that the optimal F1-Score is achieved when k=5.

When k=5, the confusion matrix is as follows:

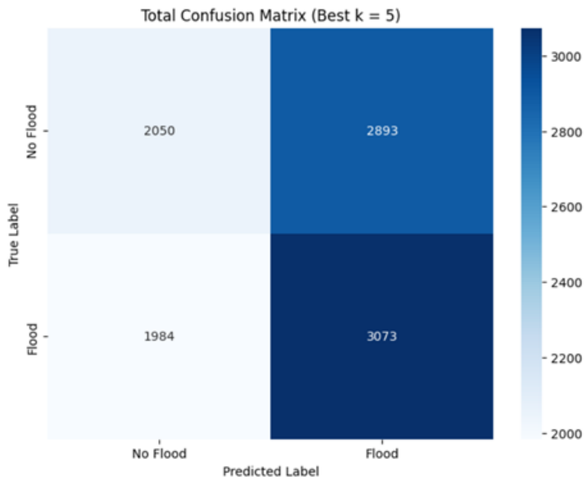

Table 3. Confusion Matrix using the K-Fold Cross-Validation technique in the Logistic Regression method.

The confusion matrix (Table 3) shows a high number of false positives (FP) and false negatives (FN) predictions, leading to unnecessary alerts and costs, and to a failure to detect actual flood events, thereby putting citizens at risk. However, the model successfully predicted 3073 flood cases (TP) and 2050 non-flood cases (TN).

Based on these results, the evaluation metrics for each dataset splitting technique are presented in Table 4:

	test_size=0.2	test_size=0.3	k-fold cross-validation (k=5)
Accuracy	49,80%	50,73%	51,23%
Error Rate	50,20%	49,27%	48,77%
Precision	50,30%	51,15%	51,50%
Recall	57,47%	57,09%	60,77%
F1-Score	53,65%	53,96%	55,75%
Specificity	41,96%	44,23%	41,47%

Table 4. Evaluation metrics of the Logistic Regression method.

The comparison of splitting techniques indicates that K-Fold Cross-Validation generally performs better, offering higher Accuracy, Recall and F1-Score, albeit at the expense of lower Specificity. In contrast, the Train-Test Split shows only slight differences in performance between different test sizes.

The same approach will be used for all machine learning methods.

3.2 K- Nearest Neighbors (KNN)

K-Fold Cross-Validation

Similar to the Logistic Regression method, the K-Fold Cross-Validation technique with k=5 provides the best overall performance in terms of average accuracy.

Moreover, KNN performs better with a larger number of neighbors, achieving optimal average accuracy at k=41 for this dataset. However, unlike the Train-Test Split results, high accuracy is also observed at lower values, such as k=5, indicating that the model can still perform effectively even with fewer neighbors.

With k=5, the confusion matrix is as follows:

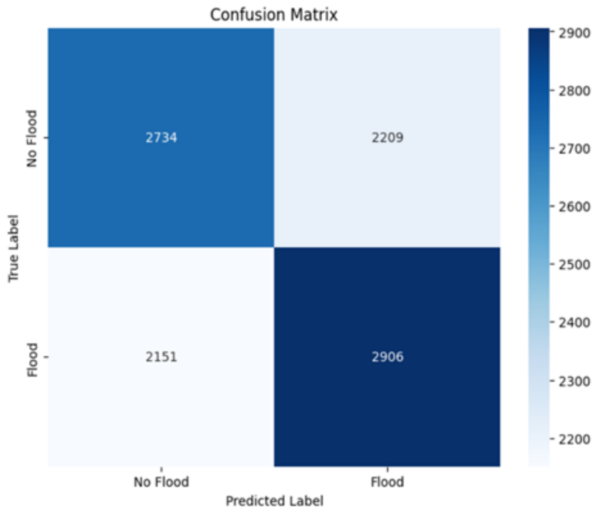

Table 5. Confusion Matrix using the K-Fold Cross-Validation technique in the K-Nearest Neighbors (KNN) method.

From Table 5, it can be observed that correct predictions outweigh incorrect ones, indicating that the model is generally reliable. Nevertheless, the rise in both false negatives and false positives cannot be ignored, as it may lead to unnecessary economic costs and represent potential risks to human life.

Also, based on the above results, the evaluation metrics that are formed depending on the dataset splitting technique are presented below:

	test_size=0.2	test_size=0.3	k-fold cross-validation (k=5)
Accuracy	51,80%	50,53%	56,40%
Error Rate	48,20%	49,47%	43,60%
Precision	53,50%	52,27%	56,81%
Recall	51,74%	48,97%	57,46%
F1-Score	52,61%	50,57%	57,14%
Specificity	51,86%	52,21%	55,31%

Table 6. Evaluation metrics of the K-Nearest Neighbors (KNN) method.

The K-Nearest Neighbors (KNN) model performs optimally with the K-Fold Cross-Validation, achieving higher accuracy and a more reliable estimate of its performance.

3.3. Naive Bayes Classifier (NBC)

K-Fold Cross-Validation

As with the other methods, using K-Fold Cross-Validation technique with k=5 achieves the highest overall performance based on average accuracy.

With k=5, the confusion matrix (Table 7) is as follows:

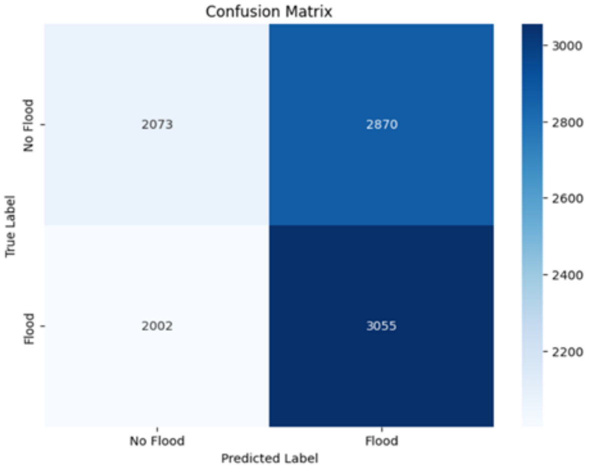

Table 7. Confusion Matrix using the K-Fold Cross-Validation technique in the Naive Bayes Classifier (NBC) method.

Although the confusion matrix (Table 7) shows more correct predictions than incorrect ones, the number of false positives and false negatives remain significant. However, the model shows improvement compared to the Train-Test Split technique.

Table 8 presents the evaluation metrics derived from the confusion matrix shown above.

	test_size=0.2	test_size=0.3	k-fold cross-validation (k=5)
Accuracy	50,90%	49,77%	51,28%
Error Rate	49,10%	50,23%	48,72%
Precision	52,26%	51,36%	51,56%
Recall	58,03%	52,32%	60,41%
F1-Score	55,00%	51,84%	55,64%
Specificity	43,27%	47,03%	41,94%

Table 8. Evaluation metrics of the Naive Bayes Classifier (NBC) method.

From Table 8, we conclude that the K-Fold Cross-Validation offers a more effective evaluation than the Train-Test Split, with higher accuracy and better detection of positive cases due to complete utilization of the data.

3.4. Support Vector Machines (SVM)

K-Fold Cross-Validation

For this method, the best overall performance, based on average accuracy, is achieved with K-Fold Cross-Validation technique at k=2.

With k=2, the confusion matrix is as follows:

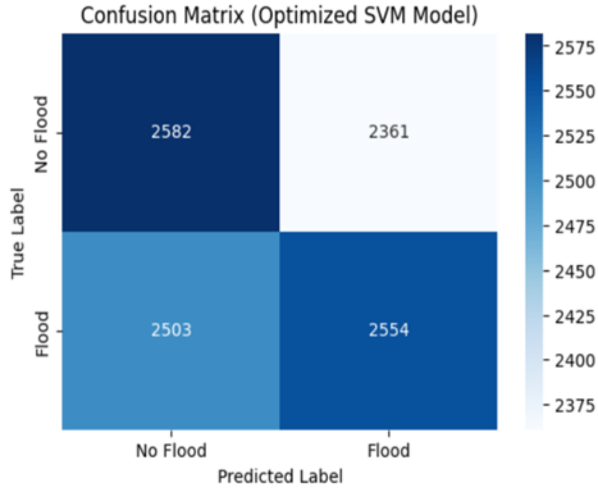

Table 9. Confusion Matrix using the K-Fold Cross-Validation technique in the Support Vector Machines (SVM) method.

The evaluation metrics are presented below (Table 10):

	test_size=0.2	test_size=0.3	k-fold cross-validation
Accuracy	50,70%	49,10%	50,88%
Error Rate	49,30%	50,90%	49,12%
Precision	52,14%	50,76%	51,44%
Recall	56,58%	49,42%	51,20%
F1-Score	54,27%	50,08%	51,32%
Specificity	44,41%	48,76%	50,56%

Table 10. Evaluation metrics of the Support Vector Machines (SVM) method.

From Table 10, we observe that with K-Fold Cross-Validation (k=2), the model achieves slightly higher accuracy (50.88%) compared to the Train-Test Split technique (50.70%). However, Train-Test Split with test_size=0.2 outperforms both techniques in terms of Precision, Recall and F1-Score.

3.5. Decision Trees

K-Fold Cross-Validation

In this method, the K-Fold Cross-Validation technique with k=12 provides the best overall performance, based on average accuracy. Furthermore, the optimal tree depth and the number of important features, which result in the highest accuracy, are 5 and 8, respectively.

With k=12, max_depth=5 and important features =8, the confusion matrix (Table 11) is as follows:

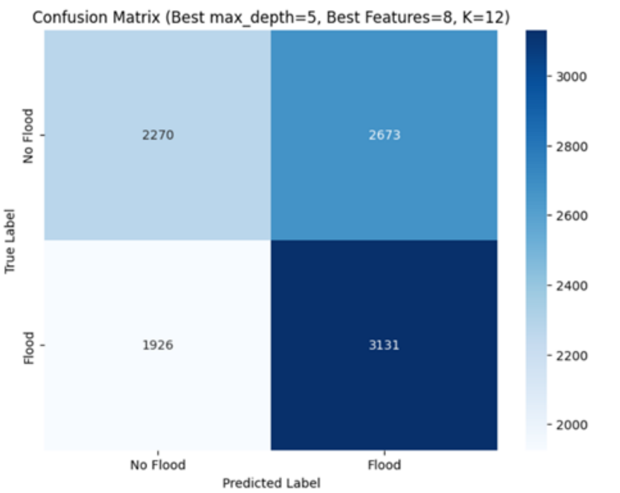

Table 11. Confusion Matrix using the K-Fold Cross-Validation technique in the Decision Trees method.

From the above matrix (Table 11), we observe that the model exhibits a bias toward flood detection, showing a high number of false positives. Despite many true positive predictions, it fails to detect non-flood cases, undermining its overall reliability.

The Decision Tree is illustrated in Fig.1:

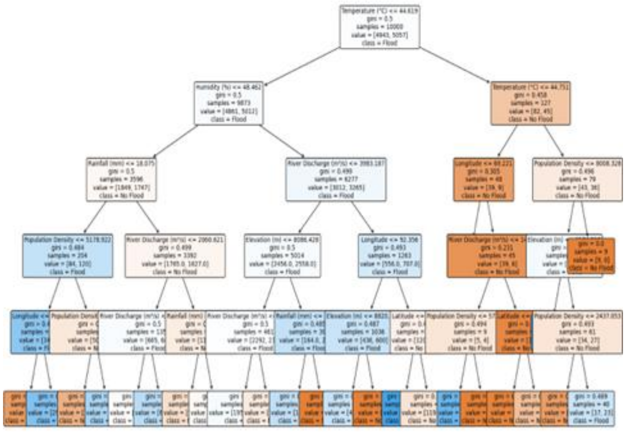

Fig. 1 Visual representation of the Decision Tree for the K-Fold Cross-Validation technique.

As shown in the decision tree (Figure 1), the most important variables for flood prediction are temperature, humidity, rainfall, river runoff, elevation, and population density. High humidity, temperature, and river runoff are associated with flood predictions, while low rainfall and population density correspond to no-flood predictions. The Gini value reflects uncertainty, and the smaller it is at the terminal nodes, the more reliable the forecast is considered.

Although the tree visualization is also useful in the Train-Test Split technique, its complexity (due to tree depth and number of important features) makes a detailed presentation unwieldy and is therefore omitted.

The evaluation metrics are shown below.

	test_size =0.2	test_size =0.3	k-fold cross-validation
Accuracy	53,15%	52,30%	54,01%
Error Rate	46,85%	47,70%	45,99%
Precision	53,67%	52,20%	53,95%
Recall	68,67%	91,16%	61,91%
F1-Score	60,25%	66,38%	57,66%
Specificity	36,54%	10,76%	45,95%

Table 12. Evaluation metrics of the Decision Trees method.

From Table 12, it is evident that while Train-Test Split with test_size=0.3 outperforms the 0.2 split in terms of Recall and F1-Score, K-Fold Cross-Validation (k=12) achieves the best overall performance, with higher Accuracy, Precision, and Specificity.

3.6 Random Forest

Train- Test Split with test_size=0.2

For the Train-Test Split with test_size=0.2, the optimal tree depth (max_depth) is 7, achieving the highest accuracy. Additionally, the features River Discharge, Water Level, and Temperature have the highest importance scores, exerting greater influence on the predictions than other variables.

The confusion matrix (Table 13) is as follows:

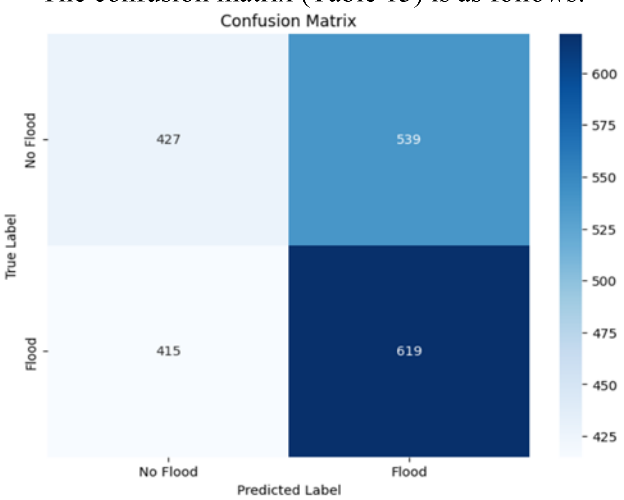

Table 13. Confusion Matrix for test_size = 0.2 using the Random Forest method.

Table 13 indicates that the model effectively identifies floods but struggles with non-flood cases, due to a high number of false positives and false negatives.

Below is a visualization of a decision tree that is part of a Random Forest using the Train-Test Split technique with test_size = 0.2 and tree depth=7 (Fig. 2).

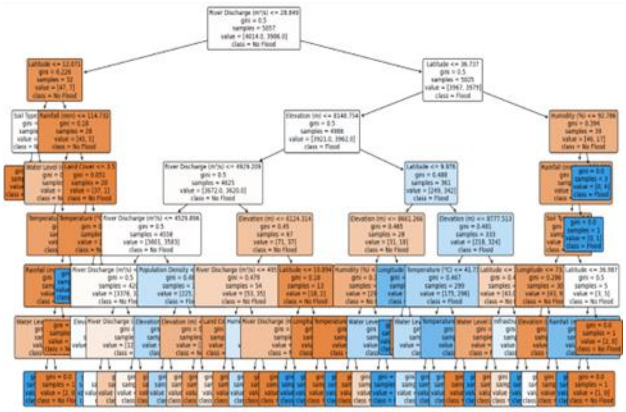

Fig. 2 Visual representation of a decision tree that is part of a Random Forest for test_size=0.2.

The decision tree highlights river discharge as the most important variable for flood prediction. Following it are geographical factors (latitude, altitude) and environmental variables (humidity, rainfall, temperature, population density), which collectively lead to the final prediction. Regions with high discharge, low latitude, and low elevation face greater flood risk, while areas with high latitude and low humidity are less vulnerable.

The performance metrics of the Random Forest model are as follows:

	test_size =0.2	test_size =0.3	k-fold cross-validation
Accuracy	52,30%	51,47%	51,66%
Error Rate	47,70%	48,53%	48,34%
Precision	53,45%	52,94%	52,10%
Recall	59,86%	54,58%	54,76%
F1-Score	56,48%	53,75%	53,39%
Specificity	44,20%	48,14%	48,49%

Table 14. Evaluation metrics of the Random Forest method.

From Table 14, we conclude that Train-Test with test_size = 0.2 achieves higher Accuracy, Precision, Recall and F1-Score, although it performs worse in Specificity. Therefore, regarding Accuracy and Recall, the Train-Test Split technique with test_size = 0.2 is better. However, considering stability and balanced metric values, K-Fold Cross-Validation is the preferable approach.

3.7 Neural Networks

In the Neural Networks method, the K-Fold Cross-Validation technique with k=10 provides the best overall performance based on average accuracy.

With k=10, the confusion matrix is as follows:

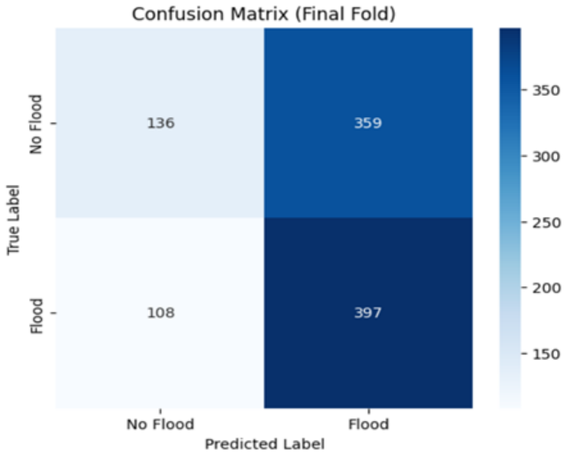

Table 15. Confusion Matrix using the K-Fold Cross-Validation technique in the Neural Networks method.

Table 15 indicates that the model tends to overpredict floods, resulting in many false positives and difficulty correctly identifying non-flood cases, which can lead to false alerts.

The evaluation metrics for this model are:

	test_size =0.2	test_size =0.3	k-fold cross-validation
Accuracy	52,60%	51,17%	53,30%
Error Rate	47,40%	48,83%	46,70%
Precision	52,07%	51,38%	52,51%
Recall	78,44%	63,88%	78,61%
F1-Score	62,59%	56,95%	62,96%
Specificity	26,19%	38,16%	27,47%

Table 16. Evaluation metrics of the Neural Networks method.

K-Fold Cross Validation technique (k=10) shows the best overall performance, mainly regarding accuracy, although low Specificity across all approaches shows difficulty in predicting non-flood events correctly.

3.8 Overall Evaluation of the Methods

In summary, we create a table of evaluation metrics for each method to compare results and identify the best overall performance.

Metric	Logistic Regression	KNN	Naïve Bayes	SVM	Decision Trees	Random Forests	Neural Networks
Accuracy	51,23%	56,40%	51,28%	50,88%	54,01%	52,30%	53,30%
Error Rate	48,77%	43,60%	48,72%	49,12%	45,99%	47,70%	46,70%
Precision	51,50%	56,81%	51,56%	51,44%	53,95%	53,45%	52,51%
Recall	60,77%	57,46%	60,41%	51,20%	61,91%	59,86%	78,61%
F1-Score	55,75%	57,14%	55,64%	51,32%	57,66%	56,48%	62,96%
Specificity	41,47%	55,31%	41,94%	50,56%	45,95%	44,20%	27,47%

Table 17. Evaluation metrics of all machine learning methods.

From Table 17, it is evident that the K-Nearest Neighbors (KNN) method with K=41 achieves the highest accuracy and outperforms in most metrics, except for Recall and F1-Score, where the Neural Networks method performs better. Consequently, K-Nearest Neighbors (KNN) is the most preferred method for flood risk prediction in India, while Neural Networks are more effective for minimizing false negatives. However, the accuracy rate of the K-Nearest Neighbors (KNN) method with K=41 (56.4%) is relatively low, likely due to the lack of critical features such as watershed boundaries, initial hydrological conditions and human interventions.

4 Conclusion

Climate change is progressively worsening over time, primarily due to rising levels of atmospheric carbon dioxide, which intensify global warming. One of its major impacts is the rise in flood events, as intense and prolonged rainfall, combined with ice melt, leads to higher water levels. Nevertheless, other factors, including urbanization, deforestation, and poor dam management, also contribute to their occurrence. Although both natural and human factors cause floods, natural ones are generally considered more hazardous because they are more difficult to manage. The consequences of floods negatively affect the economy, environment, and public health, resulting in economic instability, environmental damage, and increased risks of disease spread. To mitigate these impacts, immediate and targeted government intervention is required. Thus, flood forecasting and early warning systems are crucial for effective flood management.

In this paper, a dataset from India is used, which includes climatological, geographical, hydrological, geomorphological, and socio-economic variables to predict flood risk using machine learning methods such as Logistic Regression, K-Nearest Neighbors (KNN), Naive Bayes Classifier (NBC), Support Vector Machine (SVM), Decision Trees, Random Forests, and Neural Networks. Based on the evaluation of performance metrics for each method, it is found that the K-Nearest Neighbors method achieves the highest accuracy, at approximately 57%, making it the most effective approach for this analysis. Therefore, reducing false positives and false negatives while maximizing predictive accuracy is essential to enhance the reliability of the model.

According to the study by Debnath J. and his research team, four geomorphological features were identified as critical for reliable flood prediction in a region of India: flow accumulation, elevation, slope, and drainage density. These features were incorporated into the dataset used in their analysis, resulting in a high prediction accuracy of 92.10% [32].

Flood prediction is a complex process influenced by many factors. The main challenges includes the absence of critical features in the data (such as watershed boundaries and initial hydrological conditions), limited data quality and availability, and the use of the same model across areas with heterogeneous conditions. Furthermore, accurate flood estimation is made difficult by uncertainty in weather forecasting, particularly rainfall, which is exacerbated by unpredictable weather events associated with climate change. A further issue is the lack of recorded human interventions, including dams, water extraction and changes in the riverbed, all of which affect runoff. Lastly, sparsely populated areas often present incomplete data, making prediction even more difficult [33], [34]. The research by Emerton R. and Brunner M. confirms that significant challenges exist in predicting flood risk, which must be addressed to enhance the reliability of forecasting models.

Furthermore, in the study by Sankaranarayanan S. and his team analyzed a dataset from the Indian Water Portal, focusing on 10 regions within the Indian states of Bihar and Orissa. The parameters used for flood prediction included maximum and minimum temperature, and rainfall. By running different algorithms, it was found that the Deep Neural Network (DNN) achieved the highest accuracy, at 89.71% [35]. We can thus conclude that using another algorithm might have improved the accuracy of the current research; however, this cannot be certain due to the different dataset used.

In Table 18, there is a comparison of the current study with previous ones.

Study	Reported Accuracy (%)	Remarks
The current study (2025)	56.4%	Incorporates socio-economic variables
Debnath, et al. (2024)	92.1%	Used features such as flow accumulation and terrain slope
S. Sankaranarayanan, et al. (2020)	89.71%	Used Deep Neural Network

Table 18. Comparison of the Proposed Model with Previous Studies.

Consequently, the accuracy of flood prediction models can be enhanced by including critical features, such as flow accumulation, terrain slope, and human interventions, along with improving data quality and availability. However, achieving full accuracy remains challenging due to the inherent variability and unpredictability of natural phenomena.

For future work, the research could focus on obtaining more accurate data for variables such as rainfall volume and water level, potentially through modern remote sensing technologies, including satellites, multi-sensor systems, and high-resolution radar. In addition, combining algorithms, such as K-Nearest Neighbors (KNN) with Neural Network, could be explored to achieve higher predictive accuracy and more effective detection of flood events. Overall, the above approaches are likely to enhance the accuracy of flood prediction and, in turn, the early warning of flood events.

References:

[1] C. Change, Intergovernmental Panel on Climate Change (IPCC), CC BY, 1995, [Online]. http://energyeficiency.clima.md/files/1_Cadrul_International/2_Documente/8_IPCC/Eng/IPCC_Introduction.pdf (Accessed Date: October 13, 2024)

[2] N. Doan, et al., From Kyoto to Paris and beyond: A deep dive into the green shift. Renewable Energy, Vol. 228, 2024, pp. 1-11. doi: https://doi.org/10.1016/j.renene.2024.120675.

[3] R. Clémençon, The Two Sides of the Paris Climate Agreement: Dismal Failure or Historic Breakthrough?, The Journal of Environment & Development, Vol. 25, No. 1, 2016, pp. 3-24. doi: https://doi.org/10.1177/1070496516631362.

[4] UNFCCC, Key aspects of the Paris Agreement, [Online]. https://unfccc.int/most-requested/key-aspects-of-the-paris-agreement (Accessed Date: November 26, 2024).

[5] Europa.Eu, Συνέπειες της κλιματικής αλλαγής, [Online]. https://climate.ec.europa.eu/climate-change/consequences-climate-change_el (Accessed Date: October 15, 2024).

[6] P.H. Whitfield, Floods in future climates: a review, Journal of Flood Risk Management, 2012, pp. 336-365. doi: https://doi.org/10.1111/j.1753-318X.2012.01150.x.

[7] J. Dutta, et. al., Geomechanical Hazards related to River Hydraulics and Remedial Measures: Selected Case Studies in India, WSEAS Transactions on Fluid Mechanics, vol. 16, pp. 214-221, 2021, DOI:10.37394/232013.2021.16.20

[8] NSSL/NOAA. (2024). Flood Types, 2024, [Online].https://www.nssl.noaa.gov/education/svrwx101/floods/types/ (Accessed Date: October 22, 2024).

[9] L.J. Bracken, N.J. Cox & J. Shannon, The relationship between rainfall inputs an flood generation in south-east Spain, Hydrological Processes, Vol. 22, No. 5, 2008. doi: 10.1002/hyp.6641.

[10] A. Sendek, et al., Topographical factors related to flooding frequency promote ecosystem multifunctionality of riparian floodplains, Ecological Indicators, Vol. 132, 2021, pp. 1-11. doi: https://doi.org/10.1016/j.ecolind.2021.108312.

[11] J.Y Lee & J.S. Kim, Detecting Areas Vulnerable to Flooding Using Hydrological-Topographic Factors and Logistic Regression, Applied Sciences, Vol. 11, No. 12, 2021, pp. 1-20. doi: https://doi.org/10.3390/app11125652.

[12] E. Wohl, Streams, Encyclopedia of Inland Waters, 2009, pp. 756-765. doi: https://doi.org/10.1016/B978-012370626-3.00016-8.

[13] Z. Ö. Kurt, The effect of earthquakes on floods, Academic Studies in Engineering, 2023, pp. 47-62 [Online]. https://www.gecekitapligi.com/Webkontrol/uploads/Fck/engineering_haziran_23_yayin_1.pdf#page=53, (Accessed Date: November 2, 2024).

[14] S. Basso-Báez, et al., Unravelling the impacts to the built environment caused by floods in a river heavily perturbed by volcanic eruptions, Journal of South American Earth Sciences, Vol. 102, 2020, pp. 1-16. doi: https://doi.org/10.1016/j.jsames.2020.102655.

[15] G. Barrocu & S. Eslamian, Flood Handbook: Principles and Applications, CRC Press, 2022.

[16] B. Feng, Y. Zhang & R. Bourke, Urbanization impacts on flood risks based on urban growth data and coupled flood models, Natural Hazards, Vol. 106, 2021, pp. 613–627. doi: https://doi.org/10.1007/s11069-020-04480-0.

[17] D. P. Loucks & E. Van Beek, Water Resource Systems Planning and Management: An Introduction to Methods, Models, and Applications, Springer, 2017.

[18] A.M. Negm, G. Romanescu & M. Zeleňáková, Water Resources Management in Romania, Springer, 2020.

[19] R. Hashimoto & N. Sudo, Transmission of flood damage to the real economy and financial

intermediation: Simulation analysis using a DSGE model, Journal of Environmental Economics and Management, Vol. 128, 2024, pp. 1-20. doi: https://doi.org/10.1016/j.jeem.2024.103058.
[20] S. Boudaghpour, M. Bagheri & Z. Bagheri, Estimation of Flood Environmental Effects Using Flood Zone Mapping Techniques in Halilrood Kerman, Iran, Arab Journal of Science Engineering, Vol. 40, 2014, pp. 659–675. doi: https://doi.org/10.1007/s13369-014-1536-2.
[21] D.D Saulnier, et al., The Effect of Seasonal Floods on Health: Analysis of Six Years of National Health Data and Flood Maps, International Journal Environmental Research and Public Health, Vol. 15, No. 4, 2018, p. 665. doi: https://doi.org/10.3390/ijerph15040665.
[22] C. Stanke, et al., The effects of flooding on mental health: Outcomes and recommendations from a review of the literature, PLoS Currents, 2012. doi: 10.1371/4f9f1fa9c3cae.
[23] M. Diakakis, et al., A systematic assessment of the effects of extreme flash floods on transportation infrastructure and circulation: The example of the 2017 Mandra flood, International Journal of Disaster Risk Reduction, Vol. 47, 2020, pp. 1-17. doi: https://doi.org/10.1016/j.ijdrr.2020.101542.
[24] L. Barnard, Ten of the world's biggest flood protection projects under construction, 2024, [Online]. https://www.constructionbriefing.com/news/ten-of-the-worlds-biggest-flood-protection-projects-under-construction/8037440.article (Accessed Date: November 14, 2024).
[25] UNFCCC, Community-Based Flood Early-Warning System, India, [Online]. https://unfccc.int/climate-action/un-global-climate-action-awards/winning-projects/activity-database/community-based-flood-early-warning-system-india (Accessed Date: November 14, 2024).
[26] Vassar Labs, Flood early warning system for disaster management, 2024, [Online]. https://vassarlabs.com/flood-early-warning-system-for-disaster-management/ (Accessed Date: November 14, 2024).
[27] D. P. Patel, M. B. Dholakia, Feasible Structural and Non- Structural Measures to Minimize Effect of Flood in Lower Tapi Basin, WSEAS TRANSACTIONS on FLUID MECHANICS, Issue 3, Volume 5, July 2010
[28] C. Li, et al., Mechanisms and applications of green infrastructure practices for stormwater control: A review, Journal of Hydrology, Vol. 568, 2019, pp. 626-637. doi: https://doi.org/10.1016/j.jhydrol.2018.10.074.
[29] W. Sohn, et al., How effective are drainage systems in mitigating flood losses?, Cities, Vol. 107, 2020, pp. 1-9. doi: https://doi.org/10.1016/j.cities.2020.102917.
[30] R. Der Sarkissian, et al., Land Use Planning to Reduce Flood Risk: Opportunities, Challenges and Uncertainties in Developing Countries, Sensors, Vol. 22, No. 18, 2022, p. 6957. doi: https://doi.org/10.3390/s22186957.
[31] A. Mosavi, P. Ozturk & K. Chau, Flood Prediction Using Machine Learning Models: Literature Review, Water, Vol. 10, No. 11, 2018, p. 1536. doi: https://doi.org/10.3390/w10111536.
[32] J. Debnath, et al., Flood susceptibility assessment of the Agartala Urban Watershed, India, using Machine Learning Algorithm, Environmental Monitoring and Assessment, Vol. 196, 2024, p. 110. doi: https://doi.org/10.1007/s10661-023-12240-3.
[33] R. Emerton, et al., Continental and global scale flood forecasting systems, WIREs Water, Vol. 3, No. 3, 2016, pp. 391-418. doi: https://doi.org/10.1002/wat2.1137.
[34] M. Brunner, et al., Challenges in modeling and predicting floods and droughts: A review, WIREs Water, Vol. 8, No. 3, 2021, pp. 2-19. doi: https://doi.org/10.1002/wat2.1520.
[35] S. Sankaranarayanan, et al., Flood prediction based on weather parameters using deep learning, Journal of Water and Climate Change, Vol. 11, No. 4, 2020, pp. 1766–1783. doi: https://doi.org/10.2166/wcc.2019.321.