IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

FLOOD PREDICTION NEAR COASTAL AREAS
A Survey

[a] Saumik Idnani, [b] Vedant Punekar

[a] Student, [b] Student

Computer Science and Engineering,

Indus University, Ahmedabad, Gujarat, India

Abstract: *Floods are a devastating natural disaster that is difficult to model accurately. However, research efforts have helped to improve flood prediction models, leading to reduced risks, informed policy changes, and minimized loss of life and property damage. This paper highlights the utility of machine learning (ML) classification models in flood prediction, addressing the challenges of a complex natural disaster. Focusing on K-nearest neighbor, decision tree, and random forest models, it offers a comparative analysis for short-term and long-term flood forecasting. Moreover, the study emphasizes the importance of preprocessing techniques like min-max scaling and PCA for optimizing model performance. Through a rigorous appraisal of these techniques, the research aims to illuminate their distinctive merits and drawbacks. Leveraging the efficiency of K-nearest neighbor, the information-capturing prowess of decision trees, and the robustness of random forests, this endeavor strives to amplify the precision of flood prediction. Moreover, the integration of min-max scaling and PCA techniques aims to augment predictions through standardized feature normalization and dimensionality reduction.*

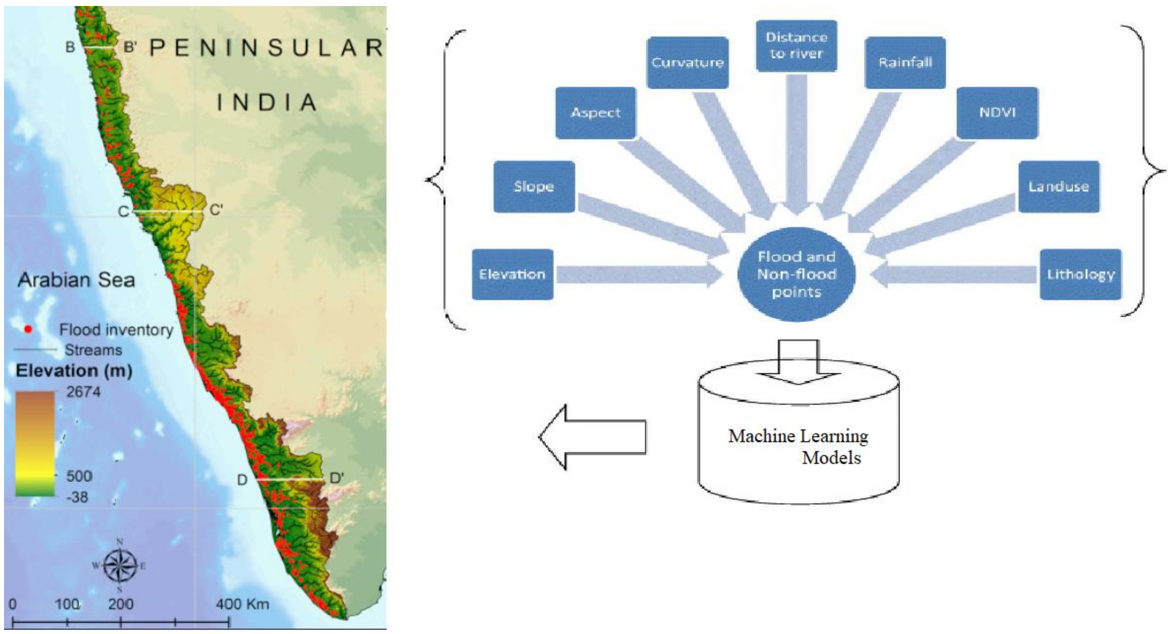
Keywords: flood prediction; flood forecasting; natural hazards & disasters; decision tree; survey; classification; data science; random forest; k-nearest neighbor; supervised learning; scaling; normalization

Introduction:

Floods are widely recognized as one of the most devastating natural disasters, causing severe damage to human life, infrastructure, agriculture, and the economy. Governments are under tremendous pressure to develop precise flood risk maps and implement sustainable flood risk management strategies that prioritize prevention, protection, and preparedness [1]. Flood prediction models play a significant role in hazard assessment and extreme event management. Accurate and reliable predictions are critical for water resource management strategies, policy analysis and recommendations, and evacuation modeling [2]. To improve the accuracy of flood predictions, it is essential to use reliable input data and integrate data from various sources such as satellite imagery, weather forecasts, and sea level measurements [3].

The research findings indicate that the application of a machine learning-based approach, which combines feature selection and classification techniques, can significantly contribute to coastal flood prediction. The proposed approach offers practical implications in disaster preparedness and risk reduction efforts, and has the potential to be used as an effective tool for coastal flood management.

The combined impact of climate change and urbanization has led to an increased risk of flooding in coastal areas, highlighting the need for effective measures to mitigate the negative impacts of floods. To address this challenge, real-time flood prediction and visualization systems have emerged as a key solution [4]. Integrating data from multiple sources is crucial for improving the accuracy and reliability of flood prediction. This includes reviewing various data sources, such as remote sensing data, weather radar data, and ground-based sensors, and highlighting their respective strengths and limitations. Real-time data assimilation is also significant in improving the precision of flood prediction models.



parameters to be defined in the model

Real-time flood prediction and visualization systems have emerged as a key solution for the increasing threat of flooding due to climate change and urbanization in coastal areas. Integration of data from various sources, such as remote sensing data, weather radar data, and ground-based sensors, is crucial to improving the accuracy and reliability of flood prediction. Additionally, real-time data assimilation is significant in increasing the precision of flood prediction models. Effective implementation of real-time flood warning and response systems requires cooperation between governmental organizations, academics, and community stakeholders to ensure the safety and well-being of coastal communities.

Implementation:

We have the dataset with historical rainfall data for different years and months, along with information about whether floods occurred. Each row in the table represents a year, and the columns provide the following information:

YEAR: The year for which the data is recorded.
JAN, FEB, MAR, ... DEC: Rainfall data for each month of the year in millimeters.
ANNUAL RAINFALL: Total annual rainfall for the year in millimeters.
FLOODS: Indicates whether floods occurred during that year ("YES" or "NO").

For example, let's take the first row:

YEAR: 1901
JAN to DEC: Rainfall data for each month (e.g., 28.7mm in January, 44.7mm in February, etc.).
ANNUAL RAINFALL: 3248.6mm (total rainfall for the year)
FLOODS: YES (indicating that floods occurred during that year)

To enhance clarity and readability, the "FLOODS" column in the data table has been dropped.This transformation ensures that the information is conveyed more explicitly, allowing for easier interpretation and analysis of the historical data.

Training and Testing

```
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.15, random_state=1)
```

This line of code uses the train_test_split function to split the input data (X) and corresponding target/output data (Y) into training and testing sets. The parameters used are as follows:

X: The input features of the dataset.
Y: The corresponding target values of the dataset.

test_size: The proportion of the dataset to include in the test split. In this case, 15% of the data will be used for testing, and 85% for training.

random_state: A seed value for random number generation. It ensures that the same random split is generated every time the code is run.

The function returns four arrays: X_train, X_test, y_train, and y_test, which represent the training and testing subsets of the input features and target values, respectively.

Classification

1. Random Forest (MinMax Scalar and PCA Scalar)
   Model Definition and Training:
   A RandomForestClassifier with 100 estimators is defined and trained using the scaled training data.
   The RandomForestClassifier is defined and trained using the PCA-transformed training data

2. K-Nearest Neighbour (MinMax Scalar and PCA Scalar)
   Model Definition and Training:
   The KNeighborsClassifier (knn) is defined with 3 neighbors and the "minkowski" distance metric.

The model is then trained on the scaled training data.The KNeighborsClassifier (knn) is defined with 3 neighbors and the "minkowski" distance metric.

The model is trained using the PCA-transformed training data.

3. Decision tree (MinMax Scalar and PCA Scalar)

Model Definition and Training:

A DecisionTreeClassifier (dt) is defined.

The features are standardized using MinMaxScaler.

The model is then trained using the scaled training data.A DecisionTreeClassifier (dt) is defined.

The model is trained using the PCA-transformed training data.

Steps

Data Scaling:

The MinMaxScaler (mms) is used to scale the features of both the training and testing data between a specific range (usually 0 to 1).

Data Standardization and PCA:

The features are first standardized using StandardScaler and then reduced to two principal components using PCA.

Model Definition and Training:

A Classifier is defined and trained using the scaled training data and the PCA-transformed training data.

Predictions:

The trained model is used to make predictions on the scaled testing data. The trained model is used to make predictions on the PCA-transformed testing data.

Evaluation and Reporting:

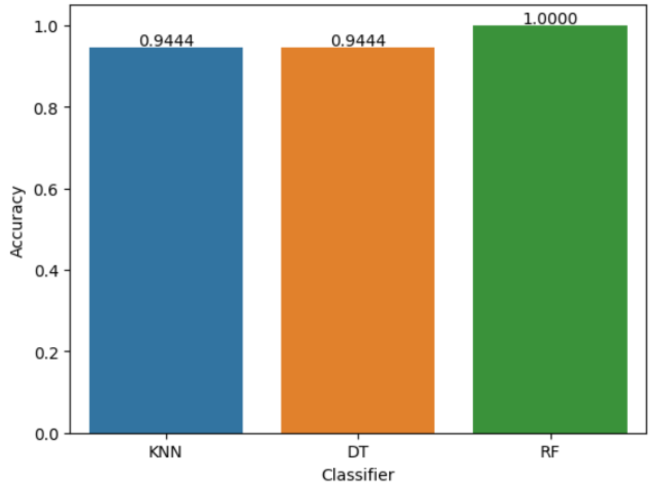
The classification report is printed to show precision, recall, and F1-score for each class ('Predicted' and 'Actual').

Confusion Matrix Visualization:

The confusion matrix is computed and displayed using ConfusionMatrixDisplay. Both normalized and raw confusion matrices are displayed using a heatmap.

Comparison

We showcase a comparison of different classifiers' performances using an ensemble of models.



the bar plot helps in quickly identifying which classifier has the highest accuracy on the given testing data

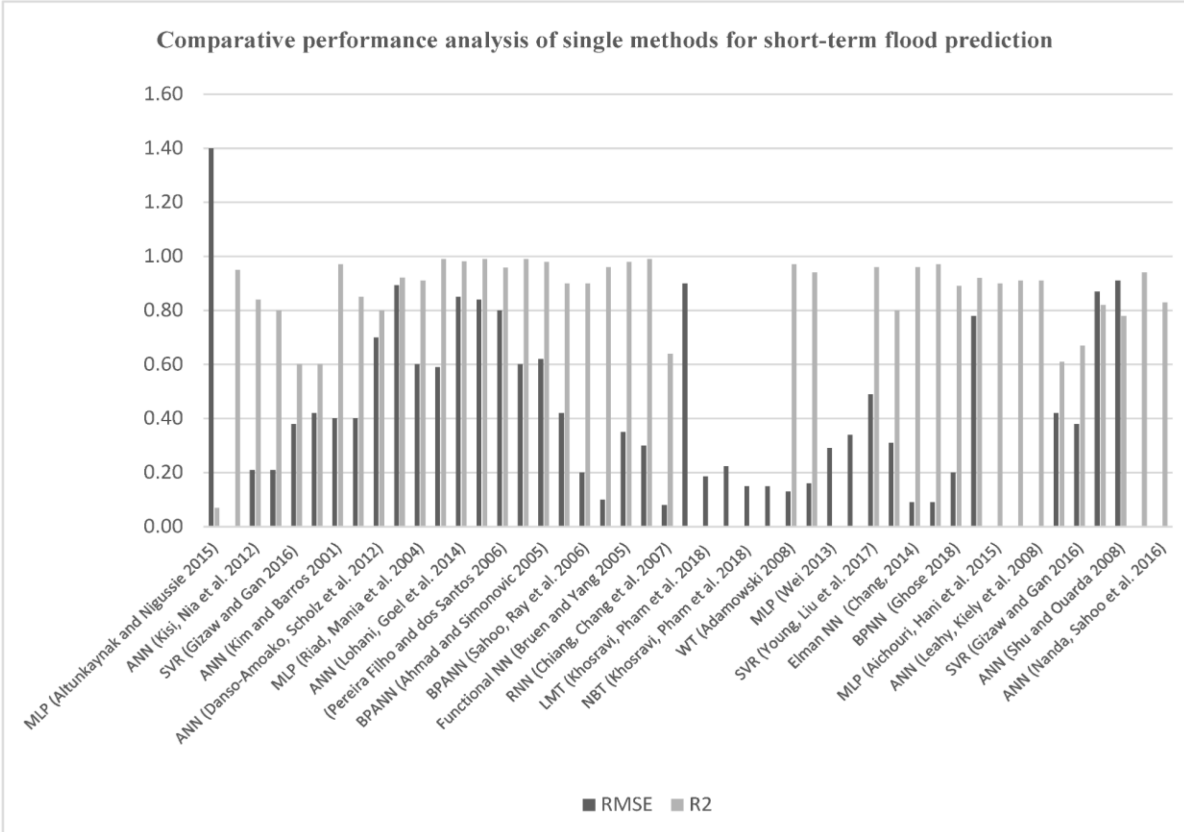
Related Work:

Coastal flooding is increasingly becoming a major challenge in urban areas, and flood prediction is a crucial aspect of addressing this problem. One proposed solution involves the use of machine learning techniques, specifically an ensemble approach that combines various models such as random forests, support vector regression, and gradient boosting regression. This approach has been shown to enhance the accuracy and reliability of flood prediction [5].

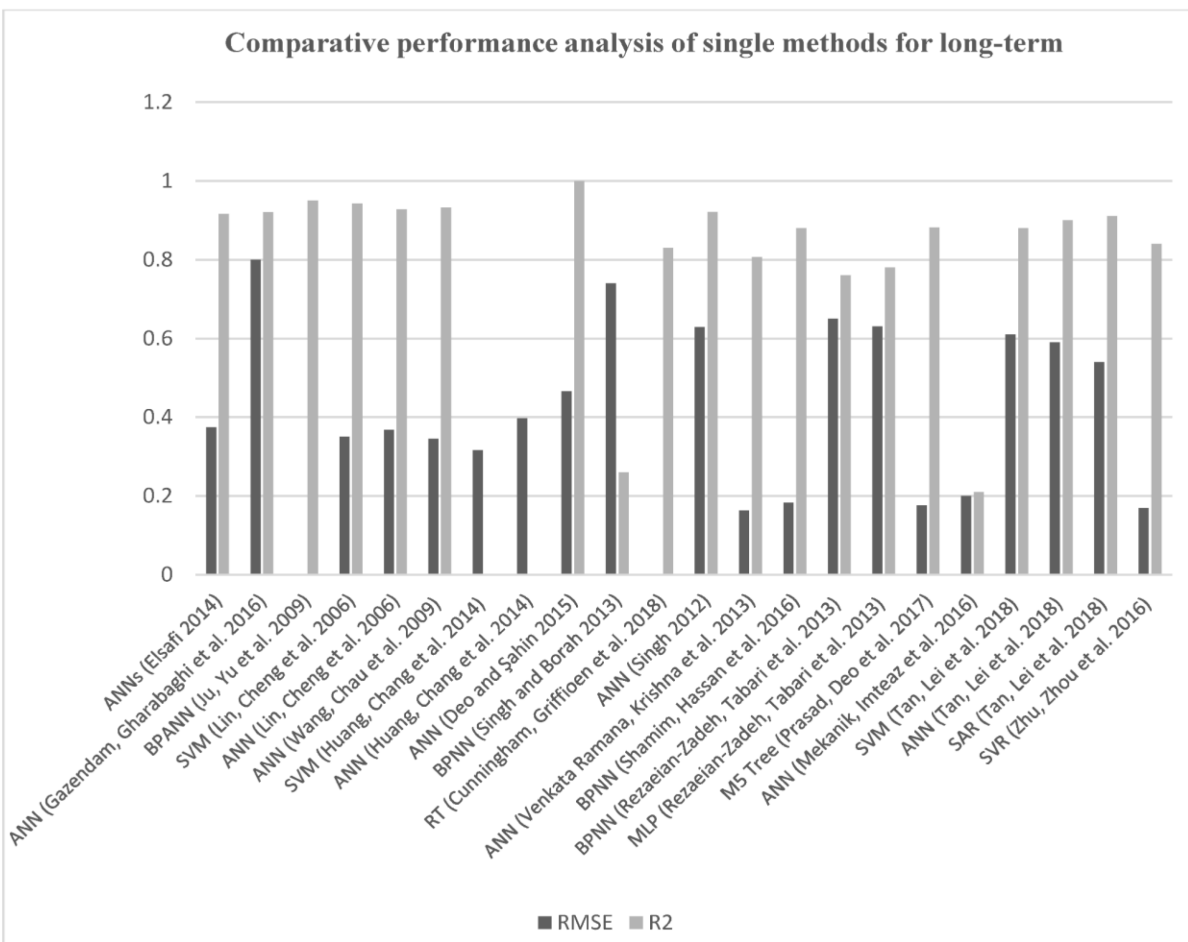
Flood prediction using classification algorithms is a method that involves using machine learning techniques to classify areas as either at risk or not at risk of flooding based on environmental and meteorological data. Popular classification algorithms used for flood prediction include decision trees, logistic regression, support vector machines, and random forests. This approach offers several advantages, such as handling large and complex datasets, automating the prediction process, and providing probabilistic predictions of flooding events. However, the accuracy of the predictions largely depends on the quality and quantity of the input data and the effectiveness of the feature selection techniques used [6].

There are various models that can be used for predicting both long-term and short-term coastal floods. The suitability of these models depends on the specific characteristics of the flood event and the data available for modeling. Here are some of the most commonly used models for predicting coastal floods:

1. Hydrodynamic models: These models simulate the physical processes that govern coastal flooding, such as storm surge and wave run-up, and provide predictions of flood extent and depth. Hydrodynamic models can be used for both short-term and long-term flood prediction, but they require detailed data on the topography, bathymetry, and boundary conditions of the coastal area being modeled.
2. Statistical models: These models use historical data on past floods and other relevant variables to predict the likelihood and severity of future floods. Statistical models can be used for both short-term and long-term prediction, but their accuracy depends on the availability and quality of historical data.
3. Artificial neural networks: These models use a complex network of interconnected nodes to learn patterns in the data and make predictions. Artificial neural networks can be used for both short-term and long-term prediction, and they can handle large amounts of data and nonlinear relationships between variables.



performance analysis graph for short term flood prediction (rmse)



performance analysis graph for long term flood prediction (rmse)

In general, hydrodynamic models are the most suitable for short-term flood prediction, while statistical and artificial neural network models are more suitable for long-term prediction.However, the best approach for predicting coastal floods depends on the specific characteristics of the flood event and the data available for modeling [7].

Using mathematical equations, hydrodynamic models replicate the physical processes that result in coastal flooding. They need information about the coastal environment, and they are calibrated using information from observed prior floods. They are useful for preparing emergency responses and developing flood management strategies since they can precisely anticipate flood extent and depth for upcoming storms. These might, however, be difficult and computationally demanding.

Using past data, statistical models for flood prediction determine the possibility of future floods. While being less complex than hydrodynamic models, they are less accurate. Although they may not be appropriate for short-term forecasting, they are useful for long-term prediction.

Artificial neural networks (ANNs) are a type of machine learning model that can learn to recognize complex patterns and relationships in data. ANNs are useful for capturing nonlinear relationships between variables, and have shown promise for flood prediction, particularly for short-term forecasting. However, they require a large amount of data for training and can be computationally intensive [8].

Datasets:

India-WRIS is a website that provides data on water resources in India, including water availability, quality, groundwater level, and rainfall data. It is a comprehensive database that can be used to understand water-related issues and to make informed decisions about water management. The website is managed by the National Remote Sensing Centre (NRSC) and provides users with access to a wide range of water-related data from different sources. It can be a useful resource for researchers, policymakers, and water resource managers.
(https://india-wris.nrsc.gov.in/)

The Central Water Commission (CWC) Water Level Data website provides real-time water level data for rivers across India. It is a valuable resource for monitoring and predicting floods in different regions of the country. The website provides data on water levels, discharge, and flood forecasts for various locations, helping to improve flood preparedness and response.It is managed by the CWC, which is the apex technical organization in the field of water resources in India.
(http://www.india-water.gov.in/ffs/real_time_water_level.htm)

Conclusion:

The use of machine learning (ML) models for flood prediction has shown great promise in improving flood risk management. Floods are one of the most catastrophic natural disasters, and their prediction is a complex task due to the numerous variables and uncertainties involved. The contribution of this paper is to demonstrate the benefits of using ML models for flood prediction and to provide insight into the most suitable models for both long-term and short-term floods.

The paper presents a comprehensive evaluation and discussion of various ML techniques for flood prediction. The performance comparison of these models helps in identifying the most effective techniques for predicting floods. The results of this study indicate that ML models can improve the accuracy of flood prediction compared to traditional methods.

The use of ML models can help in reducing the loss of life and property damage caused by floods. The insights gained from this study can be used to develop better flood risk management strategies and policies. Furthermore, the findings can also help in improving the design and implementation of early warning systems for floods.

In conclusion, this paper highlights the importance of using ML models for flood prediction and provides valuable insights into the most effective techniques for both long-term and short-term floods. The results of this study have significant implications for flood risk management, and the use of ML models can help in minimizing the impact of floods on human life and property.

References:

[1] Danso-Amoako, E.; Scholz, M.; Kalimeris, N.; Yang, Q.; Shao, J. Predicting dam failure risk for sustainable flood retention basins: A generic case study for the wider greater manchester area. *Comput. Environ. Urban Syst.* 2012, *36*, 423–433.
[2] Xie, K.; Ozbay, K.; Zhu, Y.; Yang, H. Evacuation zone modeling under climate change: A data-driven method. J. Infrastruct. Syst. 2017, 23, 04017013.
[3] "Coastal Flood Prediction using Machine Learning Techniques" by R. Jayapal, R. Rajesh and M. Ramasubramanian.
[4] "Real-time Flood Prediction and Visualization for Coastal Cities" by S. Yang, Y. Zhou and X. Wang.
[5] "An Ensemble Approach to Coastal Flood Prediction using Machine Learning" by Zhang et al. (2020)
[6] "Flood prediction using machine learning classification algorithms: A review" by Nor Azazi Zakaria et al.
[7] "A review of coastal flood forecasting models" by M. Van der Meer, J. W. T. W. Aerts, and J. C. J. Kwadijk.
[8] "Comparison of Artificial Neural Network and Time Series Models for Short-Term Flood Prediction" by B. Önöz and A. Bayazit.
[9] "Integrated coastal flood risk assessment and management" by P. Aldridge, D. Horsburgh and T. Wadey.
[10] "A Framework for Coastal Flood Risk Assessment and Prediction" by T. Wahl, M. Haigh and R.J. Nicholls.
[11] "Integrated modeling for coastal flood prediction and adaptation planning" by J. Brown, M. Zhang and J. Alberque.
[12] "Development and evaluation of an integrated flood risk assessment model for urban areas using GIS and hydrodynamic modeling" by A. Singh and A. Jain.
[13] "Machine learning for real-time coastal flooding prediction" by J. Jiang, J. Zhang, and B. Chen.
[14] "Application of artificial neural network for real-time flood forecasting in a small catchment" by S. Chatterjee and A. B. Pal.
[15] "Urban flood risk assessment under multiple sources of uncertainty" by P. Giuliani, G. Di Baldassarre, and A. Castellarin.
[16] "Flood risk assessment and management in a changing climate: A review" by A. Pathirana, S. A. Babel, and H. V. Gupta.
[17] "A comprehensive review of flood forecasting techniques and their applications" by A. K. Shrestha, S. S. Khanal, and S. Sharma.
[18] "Flood damage assessment and estimation using remote sensing and GIS: A case study of Chennai city, India" by V. G. Aswathy and M. M. Sarin.
[19] "Flood forecasting using machine learning algorithms: A review" by N. Haider, M. Ali, and A. S. Butt.
[20] "Flood vulnerability assessment using a multi-criteria decision-making approach: A case study in northern Iran" by M. Fazel, R. Ghiasi, and A. Sadeghi-Niaraki.
[21] "Flood risk assessment and mapping for small catchments using GIS and remote sensing" by S. Singh, S. S. S. Rana, and D. K. Jha.