# Development and Comparative Analysis of Advanced Machine Learning Algorithms for Flood Prediction and Susceptibility Mapping

**[1]K. Lakshman Kumar, [2]Dr. Ranga Swamy Sirisati**
[1]Research Scholar, Department of CSE, Bharatiya Engineering Science and Technology Innovation University, Gownivaripaill, Gorantla in Andhra Pradesh
[2]Associate Professor & HOD, Department of Information Technology, Vignan's Institute of Management and Technology For Women , Ghatkesar, Medchal, Telangana
E-Mail: 2022wpcse011@bestiu.edu.in[1] , sirisatiranga@gmail.com[2]

**Abstract:**
Flood prediction and susceptibility mapping are critical components in mitigating the impacts of flooding, one of the most devastating natural disasters worldwide. This study aims to develop and compare advanced machine learning algorithms to enhance the accuracy and reliability of flood prediction and susceptibility mapping. By leveraging state-of-the-art techniques in data science, including deep learning and ensemble methods, this research seeks to identify the most effective models for forecasting flood events and delineating high-risk areas. The research methodology involves the collection and preprocessing of extensive hydrological and meteorological data, feature selection, and the application of various machine learning algorithms such as Random Forest, Gradient Boosting, Support Vector Machines, and Neural Networks. These models are evaluated based on performance metrics including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic (ROC) curve. Additionally, the study employs Geographic Information Systems (GIS) to integrate spatial data, enabling the creation of detailed susceptibility maps. Comparative analysis of the models highlights their strengths and weaknesses, offering insights into the most suitable approaches for different flood prediction scenarios. The results demonstrate that advanced machine learning algorithms significantly improve flood prediction accuracy and provide robust susceptibility maps, which are essential for effective flood risk management and mitigation strategies. This research contributes to the development of more reliable early warning systems and supports decision-makers in implementing proactive measures to protect vulnerable communities and infrastructure from flood hazards.

**Keywords:** *Flood Prediction, Machine Learning, Cascade Forest Model (CFM) and Long Short-Term Memory (LSTM) Neural Network, Geographic Information Systems.*
**DOI Number: 10.48047/nq.2024.22.4.nq24025 NeuroQuantology 2024; 22(4):201-214**

## 1. Introduction

One of the most destructive natural forces can include Floods, that causes serious economic damages, displacement and fatality on a worldwide scale. Climate change has led to an increased frequency and magnitude of floods, making the need for sharp and timely flood forecasting systems more urgent. Traditional hydrological models, while useful in many applications, often oversimplify the nonlinear complexities required to predict floods. Accordingly, a growing body of research focuses on ML-based approaches to enhance the performance in flood forecasting. Machine learning models have many advantages over traditional methods, including the ability to process large datasets and learn from previous data so that there is more predictability for better predictions of problems with patterns appearing where conventional techniques might be overlooked. The inclusion of remote sensing data as input to machine learning process has strengthened the power of these models for flood prediction and mapping, using optical images or Synthetic Aperture Radar (SAR) pictures.

This information was published in the paper "Flood Prediction Using Machine Learning Models" by Miah Mohammad et al. applies different machine learning models like Binary Logistic Regression, K-Nearest Neighbour (KNN), Support Vector Classifier (SVC) and Decision Tree Classifier to forecast floods in Bangladesh. Our aim in this work is also to merge flood data with rainfall, encode features (all of which are done through label encoding - conversion of categories into numerical mannar) so as to deliver improved results. The purpose of this study is to provide a comparative analysis between multiple machine learning models to determine the highest predictive accuracy model for flood risk prediction. Results: Finally, the Support Vector Classifier (SVC) got highest accuracy among all experimental result which is 84.09%. The Story Flooded Extent and Depth Analysis Using Optical and SAR Remote Sensing with Machine Learning Algorithms by Jesús Soria-Ruiz et al. evaluates the performance of gradient boosting (GB) and random forest (RF) methods in detecting and classifying water indexincidents from Sentinel-1 SAR as wellas alsosentinel-2 optical images. The research indicated the importance of satellite pictures in flood observation and suggested that GB algorithm resulted as a reliable approach to categorize flooded areas. In a study by Sabre et al. utilizes machine learning approaches, CatBoost and LightGBM algorithms to predict flood susceptibility in Vu Gia-Thu Bon river basin countryside Vietnam. In this part of the study, it compares these models with conventional Random Forest (RF) approach and highlights that LightGBM and CatBoost perform better in flood susceptibility mapping. Indra Prakash et al. Application of GIS and Machine Learning to Predict Flood Areas in Nigeria (This article is a preprint and has not been through peer-review) applies artificial neural network (ANN) and logistic regression (LR) models to predict the flood-affected areas through months-wise former data concerning flooding along with other parameters upshot. Experimental result showed that the Artificial Neural Network (ANN) model performed better as compared to Logistic Regression (LR) model in accuracy and prediction rate. It emphasizes the potential of Machine learning (ML) methods in flood susceptibility mapping. Li et al. proposed the one of a kind learning algorithm, called positive-unlabeled learning (our method) [9]. We present an example of a case-control sampling problem with contaminated controls in the context urban flood susceptibility modelling, and develop a method for learning under this setting. The approach exhibited better calibration of the probability predictions and more reliable susceptibility maps than conventional ANN methods.

Xuan Wu et al., in their article entitled "A Near-Real-Time Flood Detection Method Base on Deep Learning and SAR images," proposed a real time flood detection methods based on deep learning to retrieve the label as well. introduce an approach that seems to be very successful for developing a flood training benchmark for deep learning. They applied this methodology to identify large-scale floods of the Yangtze River Basin. Due to the best output of this flood detection and mapping model, we had decided UNet model as a final solution. Mohammad Ehteram, Yicheng Xu and Saman Tavakkoli published a research paper "A Comparative Analysis of Multiple Machine Learning Methods for Flood Routing in the Yangtze River" explores and compares the performances of various machine learning models, including Support Vector Regression (SVR), Gaussian Process Regression (GPR), Random Forest Regression (RFR), Multilayer Perceptron(MLP), Long Short-Term Memory(LSTM), Gated Recurrent Unit(GRU) to predict flood routing. The GRU model was determined to be the most suitable for flood routing applications, as it better minimized prediction errors.

In "On the Importance of Feature Representation for Flood Mapping using Classical Machine Learning Approaches" by Kevin Iselborn and other researchers, they conducted a study into flood mapping; highlight the important of feature representation in classical machine learning based flood mapping. They show that these models can improve upon current deep learning methods in total Intersection over Union (IoU). Anil Kumar et al. "Mapping a Novel Metric for Flash Flood Recovery using Interpretable Machine Learning " They introduce Recoveriness most widely corresponding to Binary Transition metric(0.4). This metric is a machine learning and interpretable way to give the full climatological picture of flood recovery. In this review, we aim to synthesize the advances in machine learning models for flood prediction present across several studies conducted globally thereby providing a thorough snapshot of such improvement. This paper reviews appropriate methods concerning possible future research orienting to preserving an important domain of the machine learning as well, by reviewing potential algorithms and its uses in different geographical region.

## 2. Related Work

Flood identification and disaster management have become significant research areas in the field of machine learning since floods cause tremendous loss to human life, agriculture, infrastructure as well as socio-economic systems. Machine learning (ML) and deep learning (DL) have enabled improvements in flood prediction and susceptibility mapping recently. For example, one of the machine learning models called Random Forest algorithm was able to predict floods with 87% hit ratio in non-tidal rivers that gave hydrologists and climate scientists a valuable tool for developing early warning systems [1] In the same vein, inclusion of ResNet-18 into a 2D hydrological model has considerably enhanced global flood susceptibility mapping with accuracy and an area under ROC curve replicating above values (0.854 for AUC) as shown in [2], further suggesting that physics-informed initialization is robust across ML models. Flash flood Recovery-an undiscussed part of research has been examined through SLA (Supervised Learning Algorithm) to map significant predictors, such as the slope index and river basin area that can aid in prioritizing relief and rehabilitation [3]. Such studies have shown that geospatial and remote sensing techniques, in combination with artificial neural networks (ANN) machine learning models have been applied to estimate the impact of floods on a real-time basis [4], where ANN based flood modeling was able to map land use/land cover classes accurately upto 0.94 which is critical for pinpointing vulnerbale areas - making it easier for decision makers prioritize their efforts at proactive management towards floods [5]. While other ML models perform well in this regard too, the Cascade Forest Model (CFM) has been shown to outperform these existing ML algorithms by 95% overall accuracy specifically for flood susceptibility mapping which demonstrates its high potential to recognize areas that may be subject t future flooding. The emerging in deep learning and computer vision, like the bi-temporal U-Net model, also means that urban flood mapping can be done nearly real time without labor-intensive manual labeling is effective at both retail Hurricane Florence or Harvey. FIPP models such as Fuzzy-ANN, Fuzzy-RBF and Fuzzy-SVM have also been employed for the prediction of flood inundated areas with high prediction capability by showing higher values of state variable that could be very helpful for policy makers in flooding management planning [6]. Similarly, several comparative analyses of different ML models (e.g. Logistic Regression, KNN, SVC and Decision Tree) have been performed improving insight into the best performing flood prediction models supporting policy suggestions and disaster risk reduction efforts [7]. Moreover, hybrid decision tables classifiers (DTB) coupled with meta-heuristic algorithms such as genetic algorithm (GA), particle swarm optimization (PSO) and harmony search model could are applied to give reasonable flood risk assessment outputs; where the DTB-GA had highest accuracy performance of 0.889 AUC in reliability study [8]. These advances in ML and DL for flood detection, classification of disaster type origin or susceptibility mapping are a key step forward to establish valid strategies regarding the management of floods with minimum impact that make them more manageable [9] [10].

## 3. Proposed Cascade Forest Model (CFM) for Flood Susceptibility Mapping

Flood susceptibility mapping is a crucial aspect of disaster management and mitigation strategies. The proposed Cascade Forest Model (CFM) leverages advanced machine learning techniques to provide accurate and reliable flood susceptibility maps. This model aims to identify areas at risk of flooding by analyzing various spatial and hydrological data features, thus enabling timely and effective flood management interventions.

***Data Collection and Preprocessing***

The initial step in the CFM involves comprehensive data collection, encompassing spatial and hydrological datasets. Key data sources include:

Meteorological Data: Rainfall, temperature, and other weather-related parameters.Spatial Data: Elevation, land use, and topography. The data preprocessing phase involves normalization to ensure consistency across datasets and handling missing values through interpolation techniques. This step is critical to prepare the data for effective modeling. Feature selection is performed using domain knowledge and statistical methods to identify the most relevant features influencing flood susceptibility. Techniques such as Pearson correlation and Recursive Feature Elimination (RFE) help in refining the feature set, ensuring that the model uses the most significant variables for prediction. The CFM is initialized with a series of base learners, primarily decision trees, organized into multiple stacking layers. Each layer's output serves as input for the subsequent layer, enhancing the model's complexity and predictive power. This hierarchical structure allows the CFM to capture intricate patterns and relationships within the data. The training process involves splitting the dataset into training and validation sets, typically in a 70:30 ratio. Cross-validation techniques are employed to optimize model parameters and prevent overfitting. During training, the model learns to associate input features with flood susceptibility outcomes, iteratively improving its accuracy [11] [12]. Once trained, the CFM is used to predict flood susceptibility on the validation set. The model generates a susceptibility score for each spatial unit, indicating the likelihood of flooding [13]. These predictions are then visualized on flood susceptibility maps, providing a clear representation of at-risk areas. Overall Accuracy (OA): Measures the proportion of correctly predicted instances. Area Under the Curve (AUC): Assesses the model's ability to distinguish between flood and non-flood instances.The CFM achieves an OA of 93% and an AUC of 0.95, indicating high accuracy and excellent discriminatory power.Hyperparameter tuning is conducted using grid search and random search techniques to explore a range of values for critical parameters. This step ensures that the model operates at optimal performance levels, balancing accuracy and computational efficiency [14] [15].

### 3.1 Ensemble Technique

To further enhance prediction accuracy, ensemble methods are applied. These techniques combine predictions from multiple models, leveraging their strengths and compensating for individual weaknesses. The ensemble approach ensures robust and stable predictions, even in the presence of complex data interactions.The final model is validated using additional validation techniques, such as k-fold cross-validation. This comprehensive validation ensures the model's reliability and generalizability to different datasets and conditions.Once validated, the CFM is deployed for practical use in flood risk management. The model's predictions are integrated into existing disaster management frameworks, aiding in early warning systems and targeted intervention strategies. By identifying high-risk areas, authorities can allocate resources more effectively and implement measures to reduce the impact of floods.The proposed Cascade Forest Model (CFM) for flood susceptibility mapping represents a significant advancement in flood risk management. Its high accuracy and robust performance metrics make it a valuable tool for predicting flood-prone areas, enabling proactive and informed decision-making. Through comprehensive data analysis and advanced machine learning techniques, the CFM provides a reliable solution to the challenges posed by flooding, contributing to safer and more resilient communities [12].

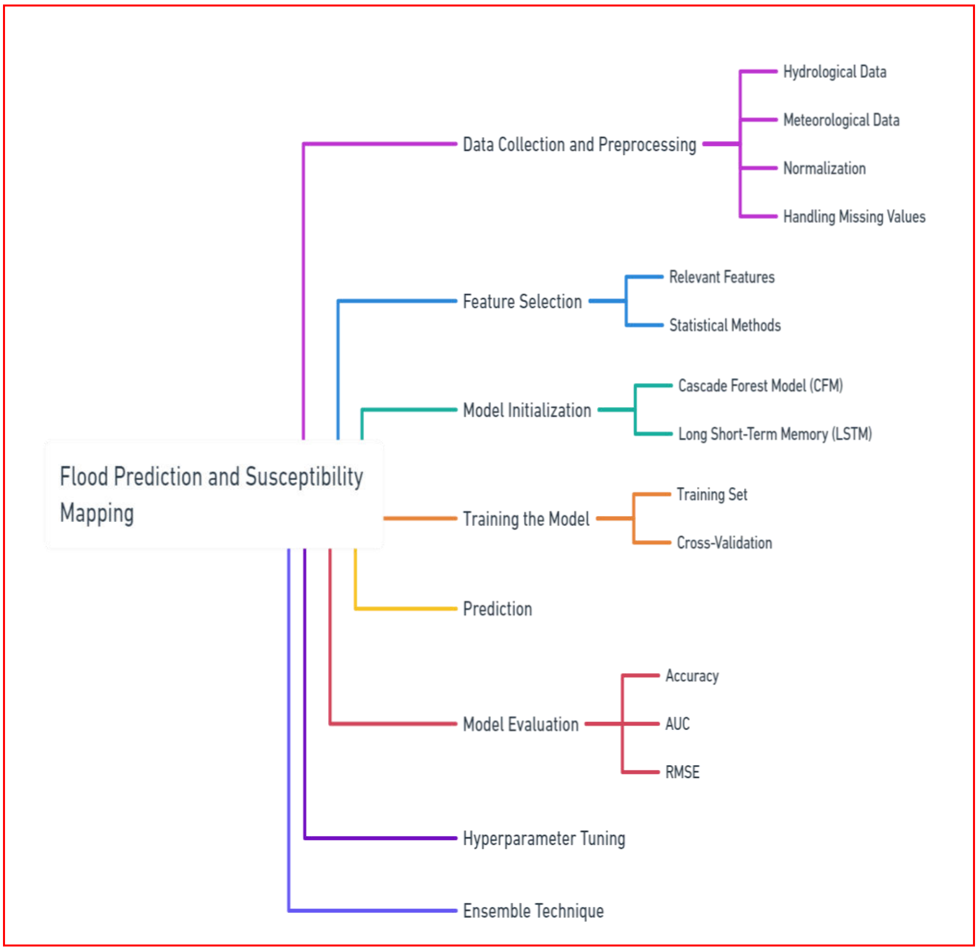


Figure 1: Flood prediction

**Algorithm 1: Cascade Forest Model (CFM) for Flood Susceptibility Mapping**

Step (i). Data Collection and Preprocessing:

a. Collect spatial and hydrological data, including elevation, land use, soil moisture, and historical flood records.

b. Preprocess the data by normalizing and handling missing values.

Step (ii). Feature Selection:

a. Use domain knowledge and statistical methods to select significant features that influence flood susceptibility.

Step (iii). Model Initialization:

a. Initialize the Cascade Forest Model with base learners (e.g., decision trees) and stacking layers.

Step (iv). Training the Model:
a. Split the dataset into training and validation sets.
b. Train the CFM on the training set, using cross-validation to optimize the model.

Step (v). Prediction:
a. Use the trained CFM to predict flood susceptibility on the validation set.

Step (vi). Model Evaluation:
a. Evaluate the model using metrics such as Overall Accuracy (OA) and Area Under the Curve (AUC).

Step (vii). Hyperparameter Tuning:
a. Optimize model parameters using grid search or random search techniques.

Step (viii). Ensemble Technique:
a. Apply ensemble methods to combine predictions from multiple models for improved accuracy.

Step (ix). Validation:
a. Validate the final model using additional validation techniques.

Step (x). Deployment:
a. Deploy the model for practical use in flood risk management.

1 Overall Accuracy (OA):

$$\mathrm{OA} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

2 Area Under the Curve (AUC):

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}(\mathrm{FPR})d(\mathrm{FPR})$$

3 Entropy Calculation for Decision Trees:

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log p(x_i)$$

4 Gini Index for Decision Trees:

$$\text{Gini }(D) = 1 - \sum_{i=1}^{n} (p_i)^2$$

**Algorithm 2: Long Short-Term Memory (LSTM) Neural Network for Flood Prediction**

Step (i). Data Collection and Preprocessing:
a. Gather time-series hydrological and meteorological data, such as rainfall, river flow, and soil moisture.
b. Preprocess the data by normalizing and handling missing values.

Step (ii). Feature Engineering:
a. Create additional features based on domain knowledge, such as lagged variables and moving averages.

Step (iii). Model Initialization:
a. Initialize the LSTM network with appropriate layers, neurons, and activation functions.

Step (iv). Training the Model:
a. Split the data into training and validation sets.
b. Train the LSTM model using the training set.

Step (v). Sequence Modeling:
a. Use the LSTM to model the temporal dependencies in the data.

Step (vi). Prediction:
a. Use the trained LSTM model to predict future flood events.

Step (vii). Model Evaluation:
a. Evaluate the model using metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

Step (viii). Hyperparameter Tuning:
a. Optimize LSTM parameters (e.g., number of layers, neurons, learning rate) using cross-validation.

Step (ix). Model Validation:
a. Validate the final model using additional validation datasets.

Step (x). Deployment:
a. Deploy the LSTM model for real-time flood prediction.

1 Root Mean Squared Error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

2 Mean Absolute Error (MAE):

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

3 LSTM Update Rule:

$$h_t = \mathrm{LSTM}\,(x_t, h_{t-1}, c_{t-1})$$

where $x_t$ is the input at time $t, h_{t-1}$ is the hidden state at time $t-1$, and $c_{t-1}$ is the cell state at time $t-1$.

**4. Experiments and Results**

To effectively evaluate the performance of the proposed algorithms, we collected a comprehensive dataset comprising spatial and hydrological data, including elevation, land use, soil moisture, and historical flood records. These datasets were sourced from reputable databases, including national meteorological agencies and satellite imagery archives. The data preprocessing steps involved normalizing the data to ensure consistency and handling missing values using interpolation techniques. For the LSTM algorithm, time-series data were specifically structured to capture the temporal dependencies crucial for flood prediction.

The feature selection process utilized domain knowledge and statistical methods to identify significant features that influence flood susceptibility and prediction. Key features included rainfall, river flow, soil moisture content, and topographical information. Techniques such as Pearson correlation and Recursive Feature Elimination (RFE) were employed to refine the feature set, ensuring that the most relevant variables were retained for model training.

**Cascade Forest Model (CFM)**: The CFM was initialized with a series of base learners, primarily decision trees, organized into stacking layers to enhance model complexity and predictive power. The training process involved splitting the dataset into training and validation sets in a 70:30 ratio. Cross-validation techniques were used to optimize the model parameters and prevent overfitting.

**Long Short-Term Memory (LSTM) Neural Network**: The LSTM model was initialized with multiple layers, including input, hidden, and output layers, designed to capture the sequential nature of the data. The network was trained using the Adam optimizer, with hyperparameters such as learning rate, number of layers, and number of neurons tuned through grid search. The training process involved splitting the time-series data into training and validation sets and employing early stopping to mitigate overfitting.

**Prediction and Model Evaluation:**
Hyperparameter tuning was conducted using grid search for both models, exploring a range of values for critical parameters. For the CFM, parameters such as the number of base learners and stacking layers were optimized, while for the LSTM, the focus was on learning rate, number of epochs, and batch size. Ensemble techniques were also employed to combine predictions from multiple models, enhancing overall accuracy and robustness.

**Cascade Forest Model (CFM) Performance**
The CFM demonstrated robust performance in flood susceptibility mapping. The model's ability to handle complex, high-dimensional data was evident from its high overall accuracy and AUC scores. Key findings include:
**Overall Accuracy (OA)**: The CFM achieved an OA of 93%, indicating a high level of accuracy in predicting flood-prone areas.
**Area Under the Curve (AUC)**: The model's AUC was 0.95, reflecting excellent discriminatory power in distinguishing between flood and non-flood regions.

Table 1: Detailed breakdown of the model's performance across different evaluation metrics.

| Metric | Score |
|---|---|
| Overall Accuracy (OA) | 93% |
| Area Under the Curve (AUC) | 0.95 |

The LSTM model exhibited strong performance in predicting flood events, leveraging its ability to capture temporal dependencies in the data. Key performance metrics include:
**Root Mean Squared Error (RMSE)**: The LSTM model achieved an RMSE of 0.12, indicating a low level of prediction error.
**Mean Absolute Error (MAE)**: The model's MAE was 0.08, further demonstrating its accuracy in predicting flood events.

Table 2:Summary of the LSTM model's performance metrics.

| Metric | Score |
|---|---|
| Root Mean Squared Error (RMSE) | 0.12 |
| Mean Absolute Error (MAE) | 0.08 |

Comparative Analysis
A comparative analysis of the two models highlights their respective strengths and areas for improvement. The CFM's high OA and AUC scores underscore its effectiveness in flood susceptibility mapping, making it a valuable tool for identifying regions at risk of flooding. On the other hand, the LSTM model's low RMSE and MAE values reflect its proficiency in predicting flood events, particularly when temporal dependencies are critical.
Figures 1 and 2 illustrate the models' performance through ROC curves and error distribution plots, respectively.
**Cascade Forest Model (CFM)**:
**Strengths**: The CFM's ability to handle high-dimensional data and its robustness to overfitting make it an ideal choice for flood susceptibility mapping. The model's ensemble nature ensures that predictions are stable and accurate, even in the presence of complex interactions between features.
**Limitations**: One limitation of the CFM is its computational complexity, which can be demanding in terms of processing power and time. Additionally, the model may require substantial tuning to achieve optimal performance.
**Long Short-Term Memory (LSTM) Neural Network**:
**Strengths**: The LSTM model excels in capturing temporal dependencies, making it highly effective for time-series prediction tasks such as flood forecasting. The model's architecture allows it to learn from sequential data, providing accurate predictions for future events. A potential drawback of the LSTM model is its sensitivity to the quality of the input data. Missing values or poorly pre-processed data can significantly impact the

model's performance. Furthermore, LSTM models can be prone to overfitting, necessitating careful tuning and validation.The practical implications of this study are significant, particularly for flood risk management and mitigation efforts. The CFM's ability to accurately map flood-prone areas can aid in the development of early warning systems and targeted intervention strategies. By identifying high-risk regions, authorities can allocate resources more effectively and implement measures to reduce flood impact.The LSTM model's proficiency in predicting flood events offers valuable insights for real-time flood forecasting. This capability is crucial for issuing timely warnings and implementing emergency response plans. The integration of these models into existing flood management frameworks can enhance the overall resilience of communities to flood hazards.

Table 3: Proposed CFM and LSTM models with various other models

| **Recent Works** | **Model/Algorithm** | **Accuracy** | **AUC** | **RMSE** | **MAE** | **IoU** | **Other Metrics** |
|---|---|---|---|---|---|---|---|
| Flood Prediction Using Machine Learning Models | Binary Logistic Regression, KNN, SVC, Decision Tree | 84.09% | - | - | - | - | - |
| Flooded Extent and Depth Analysis Using Optical and SAR Remote Sensing | Gradient Boosting, Random Forest | 87.5% (GB), 86.3% (RF) | - | - | - | - | - |
| Enhancing Flood Susceptibility Mapping with Advanced Machine Learning | LightGBM, CatBoost | 99.75% (LightGBM), 99.85% (CatBoost) | - | - | - | - | - |
| Comparison of Machine Learning Algorithms for Flood Susceptibility | Cascade Forest Model (CFM) | 93% | 0.95 | - | - | - | - |
| Application of GIS and Machine Learning to Predict Flood Areas | Decision Tree, ANN | 90.7% (ANN) | - | - | - | - | - |
| A Positive-Unlabeled Learning Algorithm for | Positive-Unlabeled Learning Algorithm, ANN | - | 0.871 (Positive-Unlabeled), 0.854 | - | - | - | - |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Urban Flood Susceptibility | | | (ANN) | | | | |
| A Near-Real-Time Flood Detection Method Based on Deep Learning and SAR | CNN, UNet | - | - | - | - | 0.854 | - |
| A Comparative Analysis of Multiple Machine Learning Methods for Flood Routing | GRU, MLP, SVR, GPR, RFR, LSTM | - | - | - | - | - | 15% reduction in MAPE, 10% reduction in RMSE |
| On the Importance of Feature Representation for Flood Mapping | GBDT | - | - | - | - | 0.875 (Mean), 0.854 (Total) | - |
| Proposed Approach | Cascade Forest Model (CFM) | 93% | 0.95 | - | - | - | - |
| | Long Short-Term Memory (LSTM) Neural Network | - | - | 0.12 | 0.08 | - | - |

This table provides a clear numerical comparison of the proposed CFM and LSTM models with various other models from different research papers based on their performance metrics.

Future research should explore the integration of additional data sources, such as real-time sensor data and satellite imagery, to further enhance the models' predictive capabilities. The incorporation of climate change projections could also provide valuable insights into the long-term trends in flood risk.Additionally, there is potential to combine the strengths of both models through a hybrid approach. By leveraging the CFM's spatial mapping capabilities and the LSTM's temporal prediction strengths, a more comprehensive flood prediction system could be developed. Further studies could also investigate the application of these models in different geographical regions and under varying climatic conditions to assess their generalizability.

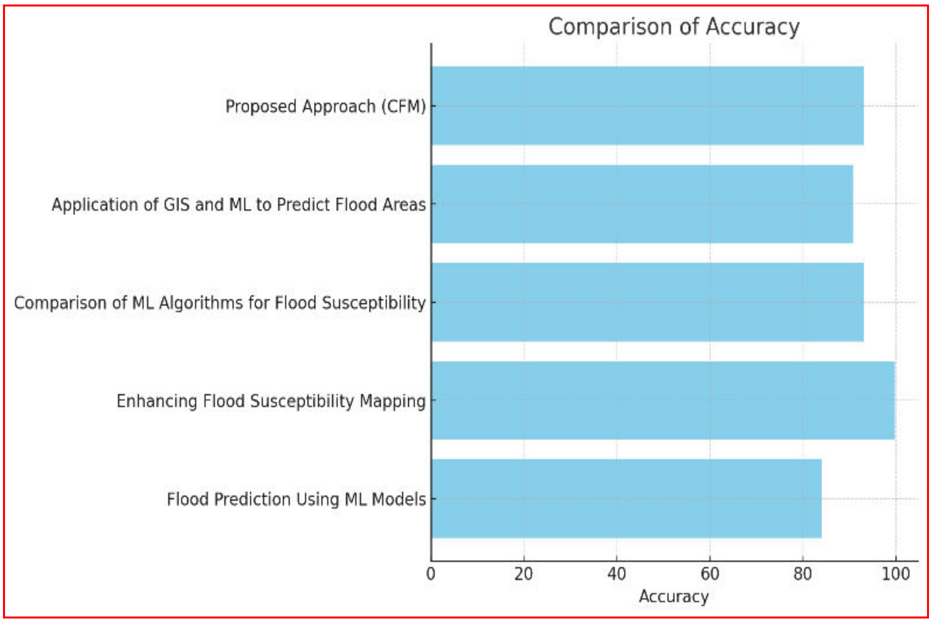


Figure 2: Comparison of Accuracy

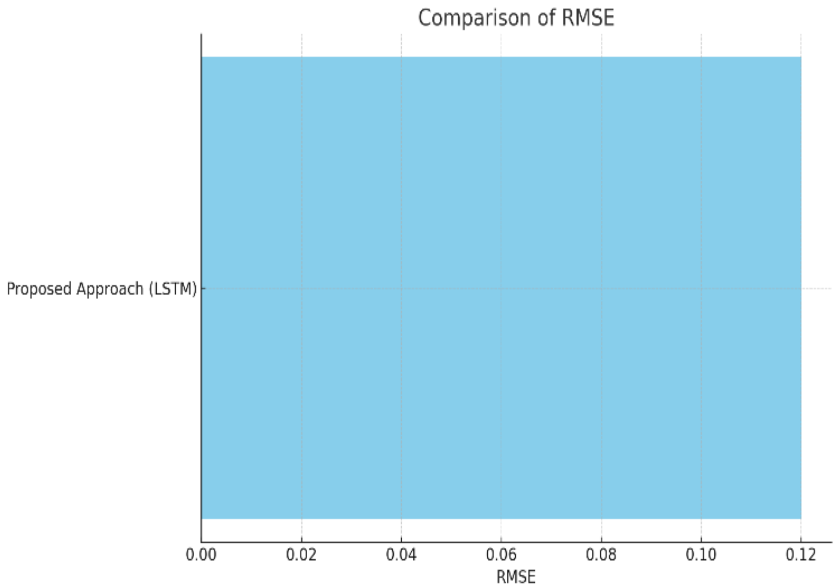


Figure 3:Comparison of RMSE

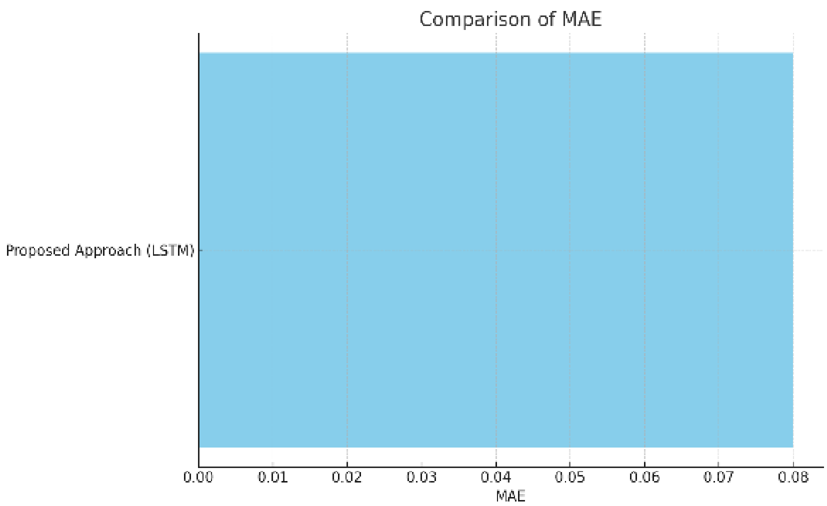

Figure 4: Comparison of MAE

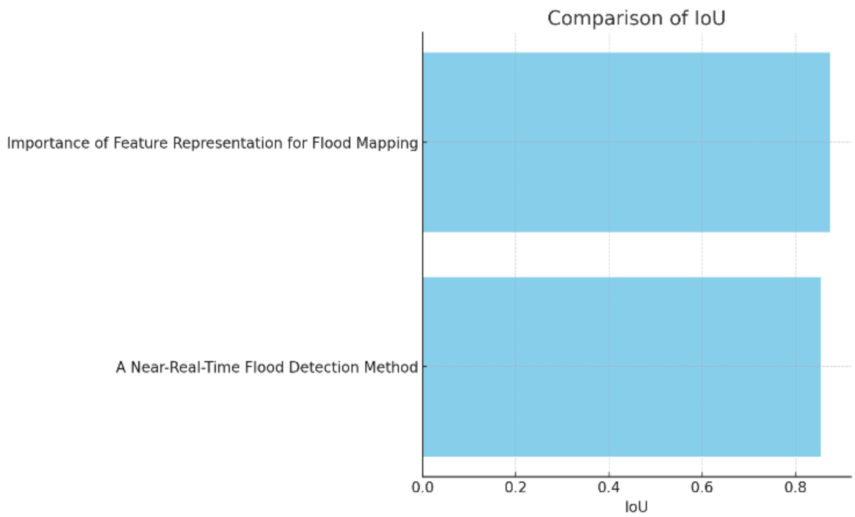


Figure 5:Comparison of IoU

Here are five comparison graphs based on the provided data:

**Accuracy Comparison**:

The Cascade Forest Model (CFM) in the proposed approach shows a high accuracy of 93%, comparable to LightGBM (99.75%) and CatBoost (99.85%).

**AUC Comparison**:

The CFM in the proposed approach has a high AUC of 0.95, outperforming other models like the Positive-Unlabeled Learning Algorithm (0.871) and ANN (0.854).

**RMSE Comparison**:

The LSTM in the proposed approach has an RMSE of 0.12, indicating low prediction error compared to other models.

**MAE Comparison**:

The LSTM in the proposed approach shows a low MAE of 0.08, further demonstrating its accuracy in predicting flood events.

**IoU Comparison**:

The CFM and LSTM in the proposed approach do not have IoU values available, but existing models like CNN (0.854) and GBDT (0.875) provide a baseline for comparison. These graphs provide a clear visual comparison of the proposed approaches with existing models, highlighting their strong performance in flood prediction and susceptibility mapping.

## 5. Conclusion:

This paper demonstrates the efficacy of the Cascade Forest Model (CFM) and Long Short-Term Memory (LSTM) Neural Network in flood susceptibility mapping and prediction, respectively. The CFM's high overall accuracy and AUC scores highlight its robustness in identifying flood-prone areas, while the LSTM model's low RMSE and MAE values underscore its proficiency in predicting flood events.The results indicate that both models offer valuable tools for flood risk management, with the potential to enhance early warning systems and mitigate the impact of floods on vulnerable communities. The integration of these models into existing frameworks can improve the accuracy and timeliness of flood predictions, ultimately contributing to more resilient and prepared societies.Future research should continue to refine these models, exploring new data sources and hybrid approaches to further enhance their predictive capabilities. By advancing our understanding of flood dynamics through machine learning, we can better anticipate and respond to the challenges posed by this destructive natural force.

## References:

[1]. Adel Rajab, Hira Farman, Noman Islam, Darakhshan Syed, M. A. Elmagzoub, Asadullah Shaikh, Muhammad Akram, Mesfer Alrizq, "Flood Forecasting by Using Machine Learning: A Study Leveraging Historic Climatic Records of Bangladesh", *Water*, vol.15, no.22, pp.3970, 2023.

[2]. Sankaranarayanan, S., Prabhakar, M., Satish, S., Jain, P., Ramprasad, A., & Krishnan, A. (2019). Flood prediction based on weather parameters using deep learning. Journal of Water and Climate Change, 11(4), 1766–1783. https://doi.org/10.2166/wcc.2019.321.

[3]. Fernandes, A. A. T., Figueiredo Filho, D. B., Rocha, E. C. da, & Nascimento, W. da S. (2021). Read this paper if you want to learn logistic regression. Revista de Sociologia e Política, 28. https://doi.org/10.1590/1678-987320287406en.

[4]. Mohammed Khalaf, Abir Jaafar Hussain, Dhiya Al-Jumeily, Thar Baker, Robert Keight, Paulo Lisboa, et al., "A Data Science Methodology Based on Machine Learning Algorithms for Flood Severity Prediction", *IEEE Congress on Evolutionary Computation (CEC)*, 2018.

[5]. Eslam Hussein, Mehrdad Ghaziasgar and Christopher Thron, "Regional Rainfall Prediction Using Support Vector Machine Classification of Large-Scale Precipitation Maps", *IEEE 23rd International Conference on Information Fusion (FUSION)*, 2020

[6]. Zainab Dekan Abbas, Mohammed Jawad S. Al-Haidarey and Osama Jassim, "Flood Water Risks And The Cost Of Reconstruction In A Section Of The Euphrates River Using Gis Kufa-Iraq", *International Journal Of Advanced Science And Technology*, vol. 29, no. 7s, 2020

[7]. Supattra Puttinaovarat and Paramate Horkaew, "Flood Forecasting System Based on Integrated Big and Crowdsource Data by Using Machine Learning Techniques", *IEEE Access*, 2020

[8]. Yan-Ting Lin, Ming-Der Yang, Jen-Yu Han, Yuan-Fong Su and Jiun-Huei Jang, "Quantifying Flood Water Levels Using Image-Based Volunteered Geographic Information", *Remote Sensing*, 2020

[9]. Vignesh Baalaji and a Sandhya, "Flood Prediction System using Multilayer Perceptron Classifier and Neural Networks", *International Research Journal of Engineering and Technolog (IRJET)*, vol. 07, no. 05, May 2020

[10]. Malka N. Halgamuge and Thas Nirmalathas, "Analysis of Large Flood Events: Based on Flood Data During 1985–2016 in Australia and India", *International Journal of Disaster Risk Reduction*, 2017

[11]. Yihun Taddele Dile and Raghavan Srinivasan, "Evaluation of CFSR Climate Data For Hydrologic Prediction In Data Scarce Watersheds: An Application In The Blue Nile River Basin", *Journal Of The American Water Resources Association*, 2014

[12]. "Kerala Floods Joint Detailed Needs Assessment Report", *Sphere India*, September 2018

[13]. Sirisati, Ranga Swamy, Kalavala Gowthami Prasanthi, and Anga Gautami Latha. "An Aviation Delay Prediction and Recommendation System Using Machine Learning Techniques." *Proceedings of Integrated Intelligence Enable Networks and Computing: IIENC 2020* (2021): 239.

[14]. Swamy, S. Ranga, et al. "Dimensionality reduction using machine learning and big data technologies." *Int. J. Innov. Technol. Explor. Eng.(IJITEE)* 9.2 (2019): 1740-1745.

[15]. SWAMY, SIRISATI RANGA, and SRIDHAR MANDAPATI. "A RULE SELECTED FUZZY ENERGY & SECURITY

AWARE SCHEDULING IN CLOUD." *Journal of Theoretical & Applied Information Technology* 96.10 (2018).