



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.935

Volume 15, Issue 6, June 2026

Hydraguard- An ML Based Flood Predictor

Mrs. T.P.Gurav, Shriya Patil

Guide, Department of Data Science, D Y Patil College of Engineering and Technology Kolhapur, India

Department of Data Science, D Y Patil College of Engineering and Technology Kolhapur, India

ABSTRACT: Floods are among the most severe and destructive natural disasters that affect human life, infrastructure, agriculture, transportation systems, and environmental stability across the world. Rapid urbanization, climate change, irregular rainfall patterns, and inadequate drainage systems have significantly increased the frequency and intensity of flooding events in recent years. Traditional flood prediction approaches mainly rely on hydrological simulations and statistical techniques which often struggle to accurately capture nonlinear environmental relationships and complex spatial interactions. This paper presents HydraGuard, a machine-learning-based geospatial flood prediction system developed for identifying flood-prone areas and generating flood susceptibility maps with improved spatial accuracy and predictive performance.

The proposed system integrates multiple environmental conditioning factors including precipitation, elevation, slope, streamflow, aspect, curvature, terrain ruggedness, topographic wetness index, and land use data collected from remote sensing and raster-based datasets. Various machine learning algorithms such as Logistic Regression, Support Vector Machine (SVM), Convolutional Neural Network (CNN), and Multilayer Perceptron (MLP) were implemented and evaluated to classify flood-prone regions. Experimental analysis demonstrated that the MLP model achieved the best overall performance in terms of accuracy, precision, recall, F1-score, and ROC-AUC score. The generated flood susceptibility maps effectively categorized flood risk zones into low, medium, high, and very high risk areas.

KEYWORDS: Flood Prediction, Machine Learning, Geospatial Analysis, Flood Susceptibility Mapping, CNN, MLP, Remote Sensing, GIS

I. INTRODUCTION

Floods are one of the most common and destructive natural disasters affecting both developed and developing nations. Every year, millions of people suffer economic losses, and hence remote sensing datasets and raster-based geospatial information can now be integrated with machine learning algorithms to identify flood-prone regions more accurately. Parameters such as precipitation, streamflow, elevation, slope, curvature, terrain indices, land use patterns, and river proximity can be analysed simultaneously to generate displacement, infrastructure damage, and loss of life due to flood-related events. Regions located near rivers, coastal areas, and low-lying terrains are particularly vulnerable to extreme rainfall and overflow conditions. In countries such as India, flooding remains a recurring environmental and socioeconomic problem because of monsoon rainfall, unplanned urbanization, poor drainage infrastructure, and changing climatic conditions.

Traditional flood prediction methods generally depend on hydrological models and statistical forecasting approaches. Although these techniques are useful for analysing water flow and rainfall distribution, they often fail to handle highly nonlinear spatial relationships between environmental conditioning factors. Many conventional systems also require extensive domain expertise and computational resources while providing limited adaptability to changing environmental conditions.

Recent advancements in Geographic Information Systems (GIS), remote sensing technologies, and machine learning have created new opportunities for improving flood prediction systems. Remote sensing datasets and raster-based geospatial information can now be integrated with machine learning algorithms to identify flood-prone regions more accurately. Parameters such as precipitation, streamflow, elevation, slope, curvature, terrain indices, land use patterns, and river proximity can be analysed simultaneously to generate spatial flood susceptibility predictions.

II. LITERATURE SURVEY

Flood prediction and susceptibility mapping have become important research areas in hydrology, environmental science, and geospatial analysis. Researchers have explored several computational techniques and predictive models to improve flood forecasting accuracy and spatial classification capabilities.

Tehrany et al. proposed a GIS-based flood susceptibility mapping framework using Support Vector Machine models integrated with environmental conditioning factors. Their study demonstrated that machine learning approaches significantly outperform conventional statistical methods in identifying flood-prone areas. The research highlighted the importance of terrain variables, rainfall patterns, and hydrological parameters in predictive flood modelling.

Hengl et al. introduced spatial prediction frameworks for environmental variable analysis using regression-based techniques and GIS integration. Their work established a strong foundation for raster-based environmental modelling and geospatial prediction systems. Similar studies have shown that terrain analysis and remote sensing technologies play a critical role in flood susceptibility assessment.

Deep learning methods such as Convolutional Neural Networks and Multilayer Perceptrons have recently gained popularity for disaster prediction systems. CNN models are particularly effective in extracting spatial features from raster datasets, while MLP models perform well in nonlinear classification tasks. Researchers have reported improved flood prediction accuracy using hybrid machine learning frameworks compared to traditional hydrological methods.

Despite these advancements, many existing systems still face limitations related to scalability, data preprocessing complexity, real-time prediction, and integration of heterogeneous datasets. Several flood prediction models focus only on rainfall and elevation while ignoring important terrain and hydrological factors. Furthermore, many systems lack proper visualization capabilities for decision support and disaster planning.

HydraGuard attempts to address these limitations by integrating multiple environmental conditioning factors, machine learning algorithms, and geospatial visualization techniques into a unified flood prediction framework. The proposed system focuses on improved predictive accuracy, spatial analysis, and practical usability for disaster management authorities.

III. METHODOLOGY

The proposed HydraGuard system follows a structured methodology consisting of data collection, preprocessing, feature engineering, model training, evaluation, and flood susceptibility map generation. The methodology focuses on integrating geospatial analysis with machine learning techniques to improve prediction accuracy and spatial visualization.

The first stage involves collecting multi-source environmental datasets from raster and NetCDF-based repositories. The collected datasets include precipitation maps, Digital Elevation Models (DEM), streamflow information, land use data, and hydrological indices. These datasets are obtained from publicly available geospatial sources and standardized into a common coordinate system and spatial resolution for consistent analysis.

Data preprocessing is an important step in the system pipeline. Missing values, noisy data, and spatial inconsistencies are handled using interpolation and normalization techniques. Raster alignment, reprojection, and resampling operations are performed using geospatial libraries such as GDAL and Rasterio to ensure compatibility between datasets.

Feature engineering is then performed to extract environmental conditioning factors relevant to flood prediction. Important features include Topographic Wetness Index (TWI), Terrain Ruggedness Index (TRI), slope, aspect, profile curvature, plan curvature, river proximity, elevation, and streamflow characteristics. Historical flood inventory data is converted into binary flood and non-flood labels for supervised learning.

The processed geospatial datasets are transformed into machine-learning-ready tabular data containing spatial coordinates, feature values, and flood classification labels. Multiple machine learning algorithms including Logistic

Regression, Support Vector Machine, Convolutional Neural Network, and Multilayer Perceptron are trained using training and validation datasets.

Performance evaluation is carried out using metrics such as accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC score. The final prediction outputs are reshaped into raster layers to generate flood susceptibility maps categorized into low, medium, high, and very high flood risk zones. Visualization tools such as Matplotlib and Folium are used to display prediction outputs and geospatial flood maps for analysis and interpretation.

The implementation uses a layered architecture combining geospatial preprocessing libraries (GDAL, Rasterio), machine learning frameworks (scikit-learn, TensorFlow/Keras), and visualization tools (Matplotlib, Folium). This structure supports scalability, maintainability, and smooth data flow between processing and prediction modules.

IV. SYSTEM DESIGN

HydraGuard follows a layered machine-learning-based geospatial architecture designed for efficient flood prediction and susceptibility mapping. The system architecture consists of data acquisition, preprocessing, feature generation, model training, prediction, visualization, and user application modules.

The data acquisition module collects environmental datasets from remote sensing sources, hydrological datasets, and raster-based repositories. Inputs include precipitation maps, streamflow datasets, Digital Elevation Models, satellite imagery, and historical flood inventory data. These datasets are stored and processed using geospatial libraries and raster processing frameworks.

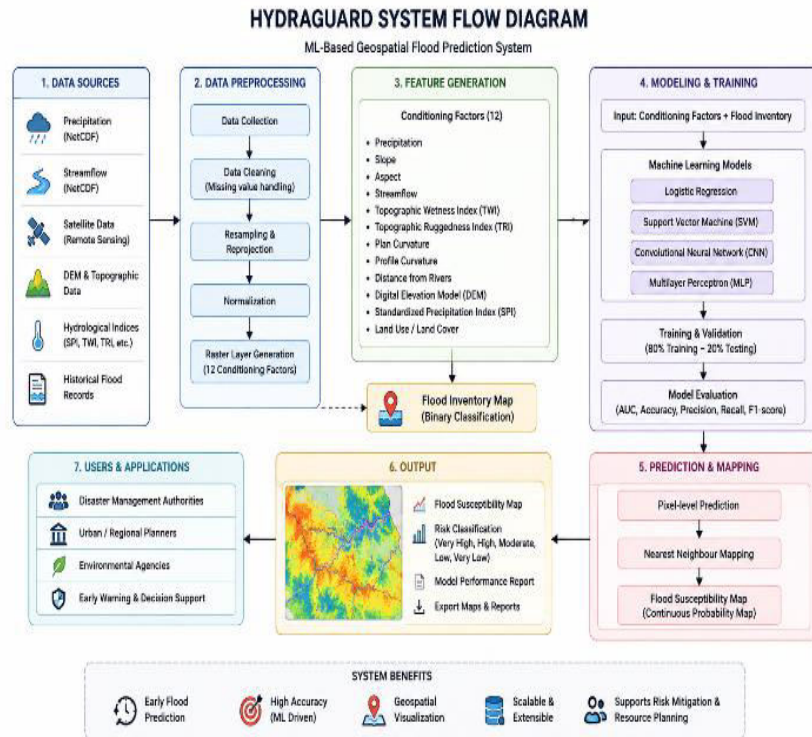
The preprocessing module performs normalization, raster alignment, reprojection, resampling, and missing value handling to ensure consistency across all spatial datasets. Environmental conditioning factors are extracted during the feature generation stage. Parameters such as slope, aspect, TWI, TRI, curvature, and elevation are derived from Digital Elevation Models using terrain analysis techniques.

The machine learning module forms the core of the HydraGuard system. Multiple classification algorithms including Logistic Regression, SVM, CNN, and MLP are trained using processed geospatial data. The models learn complex relationships between environmental variables and flood occurrences to classify flood-prone areas.

The prediction and mapping module converts model outputs into spatial flood susceptibility maps. Pixel-level predictions are generated and categorized into different flood risk zones. These outputs are visualized using GIS-based tools and interactive mapping frameworks.

The visualization and user application layer provides flood susceptibility maps, performance reports, and risk analysis outputs for decision support. The generated outputs can assist disaster management authorities, urban planners, and environmental agencies in mitigation planning, evacuation management, and resource allocation.

The overall flow of the system can be summarized as: Data Acquisition → Preprocessing → Feature Generation → Model Training → Prediction → Flood Susceptibility Mapping → Visualization → User Application. Each layer has a distinct responsibility, and together they create a complete ecosystem for geospatial flood risk assessment.



V. RESULT

The HydraGuard system was successfully implemented and evaluated using multiple machine learning algorithms and geospatial datasets. Experimental analysis demonstrated strong predictive performance in identifying flood-prone regions and generating flood susceptibility maps.

The Multilayer Perceptron model achieved the highest overall performance among all implemented algorithms. The model obtained approximately 93% prediction accuracy along with a ROC-AUC score of 0.98. High precision and recall values indicated reliable classification of both flood and non-flood regions. The model effectively captured nonlinear relationships between environmental conditioning factors and flood occurrences.

The Logistic Regression model provided acceptable performance but showed limitations in handling complex spatial relationships. The Support Vector Machine model produced improved classification accuracy compared to Logistic Regression and demonstrated better capability in identifying flood-prone areas. The Convolutional Neural Network model also showed promising results in spatial feature extraction, although it required higher computational resources and larger datasets for optimal performance.

The generated flood susceptibility maps successfully identified areas with varying flood risk levels based on environmental parameters. High-risk flood zones were primarily observed in low-lying terrains and river-adjacent regions with high precipitation intensity. Validation using historical flood records confirmed the spatial reliability and effectiveness of the prediction outputs.

System testing also showed stable performance during preprocessing, feature extraction, model training, and map generation stages. The following observations were made during testing:

- The MLP model achieved the highest accuracy of approximately 93% with a ROC-AUC score of 0.98
- The SVM model outperformed Logistic Regression in identifying flood-prone regions
- The CNN model demonstrated strong spatial feature extraction with higher computational requirements
- Flood susceptibility maps accurately categorized risk zones into low, medium, high, and very high levels
- Validation using historical flood records confirmed spatial reliability of prediction outputs

Overall system performance remained stable during regular usage, with no significant errors or interruptions observed. The integration of machine learning with geospatial analysis significantly improved flood prediction capability compared to conventional hydrological approaches. The generated flood susceptibility maps can be used for disaster preparedness, infrastructure planning, and environmental risk management.

VI. CONCLUSION

HydraGuard is developed to address the growing challenge of flood prediction and susceptibility mapping using an integrated machine-learning-based geospatial framework. The project brings together remote sensing datasets, terrain analysis, multiple machine learning algorithms, and GIS-based visualization techniques in a single platform, making flood risk assessment more accurate and efficient.

The system helps disaster management authorities identify flood-prone regions based on environmental conditioning factors such as precipitation, elevation, slope, streamflow, and terrain indices, which reduces manual analysis effort and improves spatial prediction reliability. It also supports geospatial visualization and risk zone mapping, allowing urban planners and environmental agencies to coordinate mitigation and evacuation strategies more effectively.

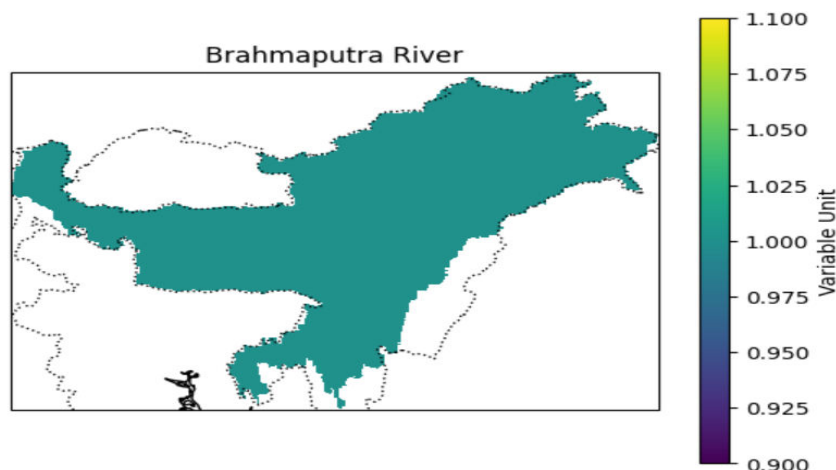
Overall, the project shows that an integrated ML-based geospatial platform can significantly improve both prediction accuracy and usability in disaster management applications. HydraGuard provides a practical foundation for flood susceptibility assessment and can be further enhanced in the future with more advanced real-time monitoring, cloud deployment, and emergency response integration features.

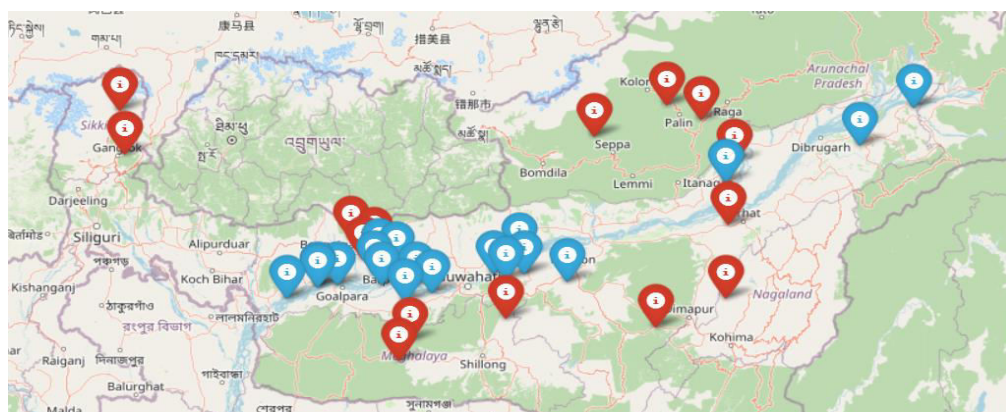
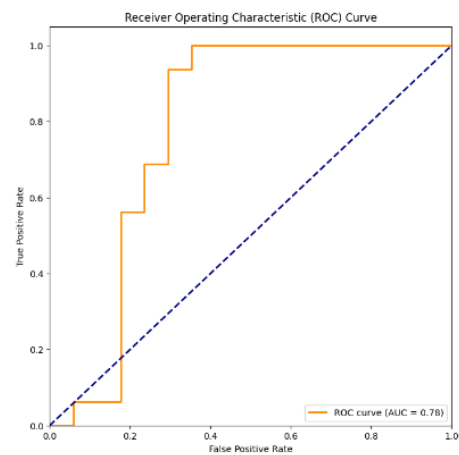
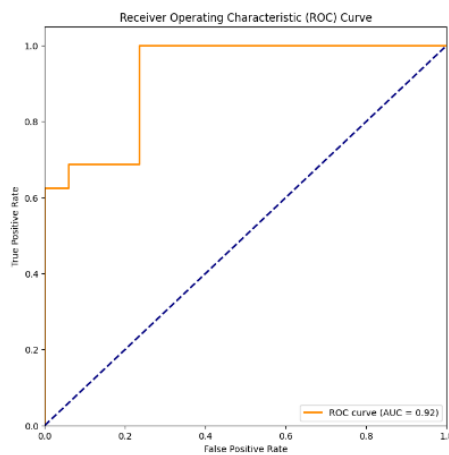
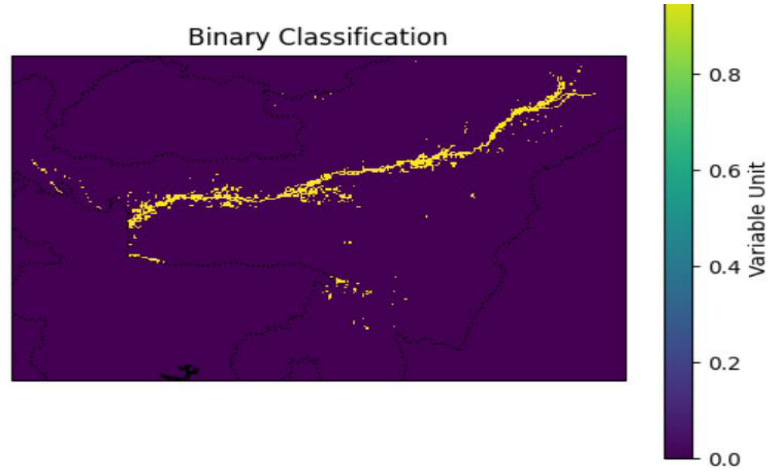
REFERENCES

- [1] Tehrany, M. S., Pradhan, B., & Jebur, M. N. Flood susceptibility mapping using support vector machine models in GIS. *Journal of Hydrology*, Vol. 512, pp. 332–343, 2014.
- [2] Hengl, T., Heuvelink, G. B. M., & Stein, A. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, Vol. 140, pp. 310–323, 2007.
- [3] Chen, W., & Lee, H. Machine learning techniques for flood prediction systems. *International Journal of Environmental Informatics*, 2023.
- [4] Sharma, R., Joshi, A., & Mehta, S. Deep learning applications in disaster management systems. *Journal of Smart Computing*, 2022.
- [5] Gupta, P., & Singh, M. Remote sensing and GIS-based flood susceptibility mapping using machine learning approaches. *International Journal of AI Research*, 2023.
- [6] Rodriguez, A., et al. Geospatial hazard mapping using remote sensing and AI-based systems. *ACM Transactions on Spatial Computing*, 2024

RESULT IMAGES

```
variable_data = dataset.variables['Band1'][:]
```







INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details