



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

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



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Urban Flooding Analysis using Deep Learning

Dr Nita Patil, Mandeep Singh, Omkar Bhogte, Sahil Bhor, Shreya Desale

Professor, Dept. of Computer Engineering, KCCEMSR, Thane, Maharashtra, India

BE, Dept. of Computer Engineering, KCCEMSR, Thane, Maharashtra, India

BE, Dept. of Computer Engineering, KCCEMSR, Thane, Maharashtra, India

BE, Dept. of Computer Engineering, KCCEMSR, Thane, Maharashtra, India

BE, Dept. of Computer Engineering, KCCEMSR, Thane, Maharashtra, India

ABSTRACT: Urban flooding poses significant challenges in metropolitan cities like Mumbai, where high population density, unplanned infrastructure, and intense rainfall events amplify flood risks. This paper presents a geospatial and machine learning-based framework for flood risk prediction and impact assessment using TIFF-based raster imagery. High-resolution spatial datasets—including .The features, combined with infrastructure and meteorological data, will be used to train multiple machine learning and deep learning models to predict flood severity, duration, human impact, infrastructure disruption, and traffic congestion. Risk zoning maps were generated to classify flood-prone areas into high, medium, and low-risk zones across Mumbai. The integration of geospatial analysis with predictive modelling provides a scalable and interpretable solution for urban flood assessment, supporting disaster preparedness, infrastructure planning, and insurance risk evaluation.

KEYWORDS: Urban flooding, geospatial analysis, TIFF imagery, hydrological modelling, Digital Elevation Model (DEM), flow accumulation, machine learning, risk zoning, spatial risk classification, QGIS.

I. INTRODUCTION

Urban flooding has emerged as one of the most critical climate-related hazards in metropolitan regions across the globe. With increasing urbanization, impervious surfaces, and extreme rainfall events, cities like Mumbai are experiencing more frequent and intense flood events. The 2005 Mumbai flood, which caused widespread damage and loss of life, underscored the urgent need for accurate flood risk prediction and proactive urban planning. Traditional flood prediction methods often rely on coarse-resolution hydrological models or historical trends that fail to capture the dynamic and spatially heterogeneous nature of urban landscapes. To address these limitations, this study proposes a scalable and data-driven approach for urban flood risk assessment using high-resolution TIFF-based raster data. By leveraging Digital Elevation Models (DEM), land use classification, flood extent, and other geospatial inputs, we derive hydrological features such as flow direction, flow accumulation, and slope. We develop multiple machine learning and deep learning models to predict key outcomes such as flood severity, water flow dynamics, impact on human life, traffic disruption, and infrastructure vulnerability. Our study focuses on Mumbai and produces risk zoning maps highlighting high, medium, and low-risk areas.

II. LITERATURE SURVEY

Urban flooding has emerged as one of the most critical climate-related hazards in metropolitan regions across the globe. With increasing urbanization, impervious surfaces, and extreme rainfall events, cities like Mumbai are experiencing more frequent and intense flood events.

1) Machine Learning (ML)-Driven Techniques

- **Motta et al. (2021)** demonstrated the application of Random Forest in urban flood prediction, achieving an accuracy of 96% and a Matthew's correlation coefficient of 0.77. However, the model exhibited a higher sensitivity, leading to a larger false-positive rate, indicating the need for further threshold adjustments.
- **Sharma et al. (2024)** evaluated seven machine learning algorithms, including XGBoost and Random Forest, for flash-flood susceptibility mapping in the Wardha River sub-basin, India. The study assessed model performance using various metrics, such as sensitivity, specificity, and area under the curve (AUC).

2) Deep Learning Methods

- **Situ et al. (2023)** proposed a hybrid feature fusion model for urban flood prediction, integrating the strengths of CNNs in processing spatial features and RNNs in analyzing temporal sequences. The model achieved high prediction accuracy and significantly improved processing speed compared to traditional methods.
- **Motta et al. (2021)** showed how to anticipate urban floods using both machine learning and geographic information systems. This project combined machine learning classifiers with GIS methodologies to create a flood prediction system for urban management.

3) Use of TIFF Images in Urban Flooding Prediction

- **Yang et al. (2020)** explored the use of remote sensing images in flood mapping by leveraging **TIFF images** obtained from satellites. They applied a **deep convolutional neural network (CNN)** to detect flood extents from high-resolution satellite images. The study demonstrated that CNNs, when trained on **TIFF images**, outperformed traditional pixel-based classification methods by achieving better accuracy in flood detection, even in complex urban environments.
- **Hao et al. (2022)** used **high-resolution SAR (Synthetic Aperture Radar)** data in **TIFF format** to detect flood-prone areas in urban regions. Their approach utilized both CNNs and traditional machine learning classifiers, and the results indicated that **TIFF images** combined with machine learning models could provide timely and precise flood predictions in regions affected by heavy rainfall.

III. METHODOLOGY

A. Data Acquisition

The study utilizes multiple high-resolution raster datasets in TIFF format covering the Mumbai Metropolitan Region. These include a **Digital Elevation Model (DEM)** at 30-meter spatial resolution, **flood extent maps** based on historical inundation events, **land use classification rasters**, **rainfall intensity layers**, and **drainage network maps**. All datasets were harmonized to a common coordinate reference system (EPSG:4326 – WGS 84) and spatially clipped to the Mumbai municipal boundary to ensure consistency in scale and extent across layers.

B. Pre-processing and Feature Extraction- Pre-processing of geospatial data was performed using QGIS 3.34 along with hydrological processing tools from SAGA GIS. The DEM was processed to remove spurious depressions using a sink-filling algorithm. Hydrological flow modelling was carried out using the D8 (Deterministic Eight Node) algorithm to compute Flow Direction, followed by the derivation of Flow Accumulation rasters. The Slope of each terrain unit was also extracted from the DEM using terrain analysis modules. All raster layers were resampled and aligned to a uniform spatial resolution and projection. Subsequently, the raster layers were sampled at the pixel level to generate a structured dataset in tabular format. Each record represented a single pixel and included attributes such as latitude, longitude, elevation, slope, flow accumulation, rainfall intensity, and land use classification.

C. Dataset Construction-

The pixel-level dataset was further enriched with urban infrastructure features such as drainage proximity and built-up density, along with corresponding flood labels derived from the flood extent raster. Each record in the final dataset was annotated with one or more target variables: a flood severity classification label (ranging from 0: No Flood to 3: High Severity), flood duration (in hours), traffic disruption index, infrastructure vulnerability score, and a human impact label indicating potential for evacuation or rescue needs. The resulting dataset comprised approximately 300,000 labeled instances, with a mix of numerical and categorical features. One-hot encoding was applied to categorical variables, and all numerical features were scaled using min-max normalization.

D. Model Architecture and Training

Multiple machine learning and deep learning models were trained to predict various flood-related outcomes. An XGBoost Classifier was implemented to perform multi-class classification of flood severity zones. A Random Forest Regressor was used to estimate flood duration, while a Support Vector Machine (SVM) was employed for binary classification of traffic disruption zones. For predicting infrastructure vulnerability, a Multilayer Perceptron (MLP) was developed using fully connected dense layers. To capture spatial context in the raster data, a Convolutional Neural Network (CNN) was trained on 64×64 pixel patches extracted from the raster stack, focusing on spatial pattern recognition for flood-prone regions. The dataset was partitioned using an 80:20 train-test split, and all models

underwent grid search-based hyperparameter tuning with 5-fold cross-validation. Evaluation metrics varied by task and included accuracy, F1-score, root mean squared error (RMSE), and receiver operating characteristic - area under curve (ROC-AUC).

E. Risk Zoning and Spatial Visualization

Model predictions were spatially reprojected and mapped back to their corresponding geographic coordinates. Based on prediction thresholds, each location was assigned a risk category: Low, Medium, or High. These classifications were then visualized as flood risk zoning maps using QGIS. Single-band pseudocolor visualization was used to represent spatial variation in risk intensity, and additional overlays were created using vector layers for roads, buildings, and drainage networks to support planning and emergency decision-making.

F. System Architecture and Deployment

The complete pipeline was integrated into a modular, real-time deployment system. A FastAPI-based backend was developed to handle model serving, data ingestion, and prediction requests. A user-facing React-based frontend was implemented to allow users to upload raster data or select predefined areas of interest. The frontend interfaces with the backend to trigger predictions and render the resulting risk zoning maps on an interactive web interface. This setup enables seamless integration of geospatial analysis with flood forecasting and decision support for both technical and non-technical stakeholders.using TensorFlow.js or ONNX.js, enabling real-time inference directly in the browser. This eliminates latency and offers end-to-end user-side processing, ensuring privacy and speed.

IV. EXPERIMENTAL RESULTS

Data Splitting and Preprocessing

The dataset used for training and testing was split in an **80:20 ratio**. Preprocessing steps included **raster alignment** to ensure consistency across multiple layers of input data, including **DEM**, **flood extent**, and **land use classification**. Additionally, **feature extraction** from these layers was performed, ensuring key variables like **elevation**, **slope**, **flow accumulation**, and **rainfall intensity** were correctly processed and normalized.

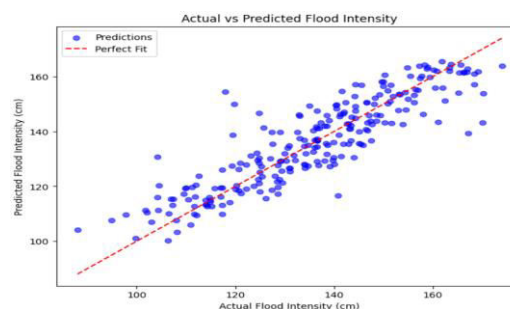
B. Training Setup and Evaluation Metrics

Each model was trained separately with hyperparameter tuning using **grid search** and validated through **5-fold cross-validation**. The following evaluation metrics were used:

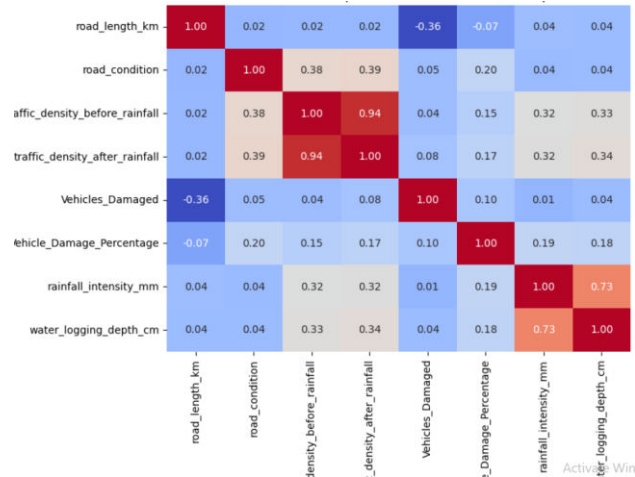
- **Flood severity classification:** Accuracy, F1-score, and Confusion Matrix
- **Flood duration prediction:** RMSE and R^2
- **Infrastructure vulnerability prediction:** RMSE and R^2

C. Model Training and Hyperparameter Tuning

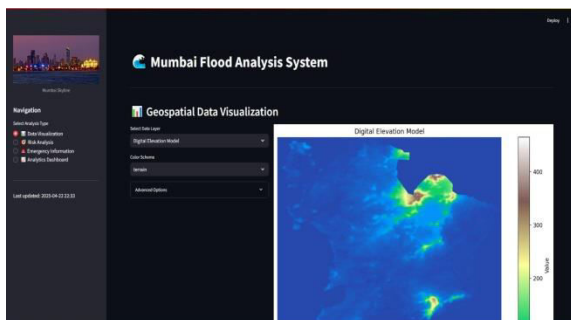
The **XGBoost Classifier** was optimized using **max_depth**, **learning_rate**, and **n_estimators** for the **flood severity classification** task. For **flood duration prediction**, the **Random Forest Regressor** was tuned with **n_estimators**, **max_features**, and **min_samples_split**. The **SVM** model for traffic disruption was trained using **C**, **gamma**, and **kernel** parameters. Finally, the **MLP** for infrastructure vulnerability and spatial flood risk prediction were tuned using varying **configurations**, **dropout rates**, and **learning rates** layer to optimize performance.



(a) Actual vs Predicted Values



(b) Heat Map for feature selection



GUI SCREENSHOTS

V. CONCLUSION

The proposed urban flood risk prediction framework offers a comprehensive, scalable, and accurate solution for assessing flood risks in urban environments, specifically in Mumbai. By leveraging machine learning models and geospatial raster data, the framework provides valuable insights into various aspects of flood risk, including flood severity, duration, traffic disruption, and infrastructure vulnerability. Compared to traditional flood prediction models like HEC-RAS and SWMM, our framework provides a more efficient and adaptable solution that can be easily scaled and integrated with real-time data. This makes it a promising tool for urban planning, disaster management, and insurance in cities worldwide.

REFERENCES

1. Kei Hiroi and Nobuo Kawaguchi et al "FloodEye: Real-time flash flood prediction system for urban complex water flow" 2017.
2. Christoph Hardegen, Benedikt Pfulb, Sebastian Rieger, Alexander Gepperth, and Sven ReiBmann "Flow-based Throughput Prediction using Deep Learning and Real-World Network Traffic" year 2019.
3. Charalampos Karyotis, Tomasz Maniak, Faiyaz Doctor, Rahat Iqbal, Vasile Palade, and Raymond Tang "Research and Discussion on Image Recognition and Classification Algorithm Based on Deep Learning" year 2021.
4. Pooja Gupta , "Rainfall Time Series Data", " Kaggle, 2022. [Online]. Available: <https://www.kaggle.com/datasets/poojag718/rainfall-timeseries-data> [Accessed: Aug, 2024].
5. "Working with TIFF files in Python using Rasterio," Rasterio Documentation. [Online]. Available:

<https://rasterio.readthedocs.io/en/latest/>. [Accessed: Sept. 2024].

6. "Sentinel-1 Toolbox - SNAP," European Space Agency (ESA). [Online]. Available:
<https://step.esa.int/main/toolboxes/snap/>. [Accessed: Sept. 2024].
7. "TIFF File Format Specification," Adobe Developer Center. [Online]. Available:
<https://www.adobe.io/open/standards/TIFF.html>. [Accessed: Sept. 2024].



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