

Flood Susceptibility Mapping in Kolhapur City

Pranav A. Bhosale, Pranav P. Mane, Prachi P. Jadkar, Purva V. Bramhadande, Anjali U. Urunkar

Students, Department of Computer Science and Engineering (Data Science)

D. Y. Patil College of Engineering and Technology

Kolhapur, India

bhosalepranav2021@gmail.com

Dr. G. V. Patil

Prof., Department of CSE (Data Science)

D. Y. Patil College of Engineering and Technology

Kolhapur, India

pganeshv@gmail.com

Abstract—Flooding remains a dominant environmental hazard affecting urban infrastructure, and Kolhapur city is critically vulnerable due to concentrated monsoon precipitation and the topological complexity of the Panchganga River basin [1], [2]. Recurring inundation continues to paralyze transportation networks and agricultural productivity. Preemptive identification of spatial vulnerability is mathematically necessary for targeted disaster mitigation [8]. This research presents a comprehensive comparative evaluation of flood susceptibility utilizing three distinct predictive architectures: the traditional Frequency Ratio Model (FRM), Support Vector Machine (SVM), and Random Forest (RF) [4], [9]. The methodological framework incorporated ten high-resolution geo-environmental conditioning factors: elevation, slope, rainfall, land use/land cover, river proximity, Topographic Wetness Index (TWI), Stream Power Index (SPI), Terrain Ruggedness Index (TRI), plan curvature, and aspect [7], [10]. To enforce strict comparative parity, all algorithms processed identical standardized training and validation matrices compiled via Google Earth Engine and QGIS. Model performance was quantified using the Receiver Operating Characteristic Area Under Curve (ROC-AUC), classification accuracy, and confusion matrix analytics. Empirical results demonstrate that the Random Forest ensemble significantly outperformed both the statistical baseline and the SVM, achieving a predictive accuracy of 90.28% and an AUC of 0.93. Feature importance analytics confirmed distance to drainage and absolute elevation as the primary drivers of localized inundation. The resulting susceptibility mapping stratifies the district into five distinct risk gradients, delivering a highly reliable spatial intelligence tool for municipal zoning and emergency protocol development.

Index Terms—Flood susceptibility, Random Forest, Frequency Ratio Model, Support Vector Machine, Spatial Analysis, GIS, ROC-AUC, Optimization.

I. INTRODUCTION

The increasing frequency of severe urban flooding has elevated regional hydrological management into a priority area of spatial data science [5], [8]. Within Maharashtra, India, Kolhapur city functions as a highly susceptible geographic basin. Flanked by the Panchganga River and subjected to the intense climatic forcing of the Western Ghats, the district absorbs massive volumes of monsoonal precipitation [1], [2]. Exceptional flood events documented in 2005, 2019, and 2021

triggered vast economic losses, displaced local populations, and severely compromised civic infrastructure [1], [2]. As accelerating, largely unmonitored urban expansion continues to replace permeable soils with impermeable concrete, the natural drainage capacity is continuously altered, making the spatial prediction of future flood zones a complex modeling challenge [13].

Historically, hydrological vulnerability mapping relied heavily on bivariate statistical frameworks, with the Frequency Ratio Model (FRM) serving as the academic standard [7], [10]. FRM quantifies the spatial relationship between documented historical inundations and discrete environmental parameters. It remains widely utilized due to its computational simplicity and high interpretability within standard Geographic Information Systems (GIS) [3], [14]. However, the physical reality of a flood involves highly non-linear, multi-dimensional interactions [9]. For example, the threshold at which soil saturation leads to surface runoff is dictated simultaneously by slope angle, vegetation cover, and localized terrain roughness. Statistical approaches fundamentally assume a degree of linearity and variable independence that simply does not exist in complex topography, thereby limiting their predictive accuracy in highly heterogeneous environments [8].

In response to these limitations, contemporary hydrological modeling has increasingly integrated advanced machine learning algorithms [4], [12]. Architectures such as Random Forest (RF) and Support Vector Machine (SVM) possess the mathematical capacity to ingest high-dimensional spatial data and map non-linear environmental interactions without relying on restrictive statistical assumptions [4], [5]. While the theoretical advantages of machine learning are well documented, a pervasive flaw exists in current literature regarding comparative evaluation. Researchers frequently test different algorithms using varying spatial resolutions, differing data resampling techniques, or isolated training sets [6]. This introduces severe methodological variance, making it difficult to ascertain whether a model succeeded due to its underlying algorithmic efficiency or a superior preprocessing pipeline

[11].

This research addresses this gap by establishing a strictly controlled comparative environment. The primary objective is to execute a rigorous evaluation of FRM, SVM, and RF under identical experimental constraints. By standardizing the geographical bounding box, preprocessing normalization, and the historical validation dataset within Google Earth Engine, the algorithmic performance of each model is isolated and quantified. This framework definitively identifies the most mathematically sound methodology for mapping flood susceptibility in Kolhapur, culminating in an actionable hazard map designed to guide municipal infrastructure planning [11].

II. LITERATURE REVIEW

Flood susceptibility mapping has transitioned rapidly from qualitative geographical assessments to rigorous quantitative data science. Early methodologies utilized basic multi-criteria decision overlays. As spatial computational tools evolved, statistical modeling became the default mechanism. For instance, Lee and Pradhan [14] established foundational frameworks for probabilistic hazard mapping using the Frequency Ratio Model. Building on this, Powar and Panhalkar [1] demonstrated the efficacy of the FRM specifically within the Kolhapur district, proving that historical spatial correlations could accurately identify primary inundation zones. Corroborating studies across various topographies have established that FRM delivers acceptable baseline metrics when supplied with high-resolution digital elevation models [2], [8].

However, as researchers expanded their parameter lists to include complex hydrological derivatives such as the Stream Power Index and Topographic Wetness Index the structural limitations of bivariate statistics became problematic. Tehrany et al. [9] and Arabameri et al. [3] highlighted that traditional models struggle to process multi-collinearity, often failing to appropriately weigh overlapping environmental influences. This mathematical bottleneck catalyzed the adoption of machine learning. Early transitions into artificial intelligence, such as the artificial neural network models developed by Kia et al. [15], proved that algorithmic frameworks could map non-linear hydrological behaviors far better than linear statistics.

Following this, Support Vector Machines gained significant traction for their margin-based classification mechanics [12]. By utilizing kernel functions to project spatial data into higher-dimensional planes, SVMs successfully separate flood-prone pixels from safe terrain even when the underlying data appears chaotic. Comprehensive reviews, such as the one conducted by Mosavi et al. [10], confirmed SVM as a leading algorithm for hydrological forecasting. More recently, ensemble tree-based algorithms, specifically Random Forest, have dominated spatial modeling [11], [13]. Khosravi et al. [5] emphasized RF's inherent resistance to training-data overfitting and its critical capability to calculate feature importance.

Despite this progression, academic literature concerning the Panchganga basin remains highly segmented. While comparative analyses of machine learning methods exist globally [4], [6], and hybrid deep-learning models are emerging [7],

there is a distinct absence of unified, standardized comparative research that evaluates both traditional statistics and ensemble machine learning on an identical spatial matrix for Kolhapur city.

III. STUDY AREA AND DATA PREPARATION

The integrity of a spatial predictive model relies entirely on precise geographic scoping and meticulous data curation [7]. The structural architecture of this study was divided into discrete computational phases: data ingestion, inventory synthesis, and spatial standardization.

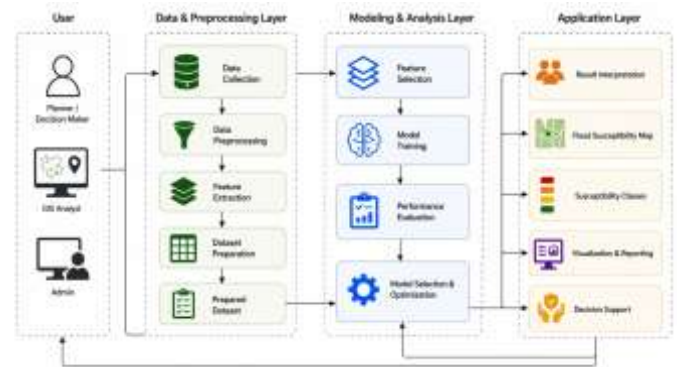


Fig. 1. Proposed system architecture for flood susceptibility analysis.

A. Study Area Description

The computational boundary encompasses Kolhapur city and the surrounding hydrological catchment. Located in south-western Maharashtra, the topography is characterized by a low-gradient central basin bordered by undulating elevations [1], [2]. The dominant hydrological vector is the Panchganga River system, sustained by a dense tributary network. During the active monsoon season, precipitation intensity frequently exceeds the natural soil infiltration rate. The resulting surface runoff converges rapidly toward the low-elevation urban center, triggering localized flash floods and broad riverine inundation [2], [7].

B. Flood Inventory Compilation

Supervised learning models require accurate dependent variables (ground truth) to optimize their internal weights [4]. A historical flood inventory was assembled utilizing verified field reports and geographic records. Exactly 178 precise coordinates corresponding to empirical flood events were digitized as point vector data.

To train the SVM and RF classifiers to establish decision boundaries, negative samples were required. Non-flood points were generated synthetically through stratified random sampling. These points were extracted strictly from high-elevation, steep-gradient zones located geographically distant from active drainage channels, ensuring a balanced, binary classification target.

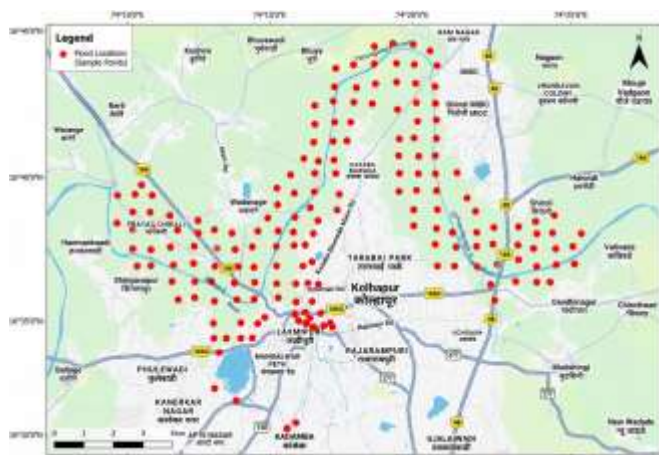


Fig. 2. Location map of flood inventory points in Kolhapur city.

C. Selection of Flood-Conditioning Factors

Inundation is governed by the physical interaction of topography, hydrology, and meteorological forcing. Based on geospatial availability and domain relevance, ten independent conditioning factors were processed as raster matrices to comprehensively map the environment [1], [10].

1. Elevation: Extracted from high-resolution Digital Elevation Models (DEM), absolute elevation fundamentally controls gravity-driven water movement [1], [7]. Lower elevations naturally act as accumulation sinks for runoff originating from higher terrain, making them intrinsically vulnerable to standing water.

2. Slope: This topographical factor dictates the velocity of surface water [2]. Steep gradients accelerate runoff and severely reduce the time available for soil infiltration. This rapid transit pushes massive volumes of water toward flat, low-slope urban areas where drainage efficiency drops, causing immediate inundation.

3. Rainfall: As the primary climatic trigger, heavy monsoon precipitation overwhelms the local hydrological cycle [5]. Variations in spatial rainfall intensity directly correlate with flash flood hotspots, as excess water immediately exceeds the infiltration capacity of the ground.

4. Distance from River: The Euclidean proximity to the Panchganga River and its major tributaries is a critical hydrological metric [2]. Areas closest to the channels are mathematically the first to experience overflow when the river breaches its carrying capacity.

5. Topographic Wetness Index (TWI): TWI quantifies topographical control over hydrological processes [9]. It mathematically represents a pixel's physical tendency to accumulate moisture based on its specific upstream contributing area and the localized slope gradient, identifying zones where water naturally pools.

6. Stream Power Index (SPI): SPI measures the erosive kinetic energy of flowing surface water. High SPI values indicate areas where concentrated surface runoff possesses

significant energy, which directly contributes to rapid flood onset and soil erosion during storm events [10].

7. Terrain Ruggedness Index (TRI): TRI calculates the localized morphological variance of the terrain. Highly rugged, irregular terrain disperses water flow in multiple directions, whereas smooth, uniform terrain allows for massive, uninterrupted pooling of floodwaters.

8. Plan Curvature: This measures the horizontal curvature of the earth's surface [12]. Concave terrain profiles physically force surface runoff to converge into specific choke points, rapidly increasing localized flood depth, whereas convex surfaces safely disperse the flow radially.

9. Aspect: Aspect identifies the compass direction a specific slope faces. It influences local microclimates, vegetation density, and soil moisture evaporation rates, subtly affecting how long an area remains saturated prior to a major rain event.

10. Land Use/Land Cover (LULC): Extracted via multi-spectral satellite imagery, LULC differentiates between permeable natural surfaces (like agriculture and forests) and highly impermeable urban infrastructure [1], [7]. High urbanization drastically reduces infiltration, causing a near-instantaneous spike in peak surface runoff.

TABLE I
FLOOD-CONDITIONING FACTORS USED FOR FLOOD SUSCEPTIBILITY ANALYSIS

Factor	Category	Influence on Flooding
Elevation	Topographical	Controls water accumulation and drainage flow
Slope	Topographical	Influences runoff velocity and infiltration.
Rainfall	Climatic	Increases surface runoff and flood intensity
Distance from River	Hydrological	Indicates proximity to flood-prone regions
TWI	Hydrological	Represents moisture accumulation tendency
SPI	Hydrological	Measures erosive power of flowing water.
TRI	Topographical	Represents terrain roughness and variation
Plan Curvature	Terrain Parameter	Influences flow convergence and divergence
Aspect	Terrain Parameter	Affects slope orientation and runoff behavior
LULC	Environmental	Represents surface cover and infiltration pattern

IV. METHODOLOGY AND ALGORITHMIC FORMULATIONS

Transitioning raw geospatial arrays into a format suitable for algorithmic ingestion requires rigorous preprocessing and mathematical formulation [8], [12].

A. Data Preprocessing and Standardization

Raw geographic datasets exhibit differing spatial resolutions and projections. Uncorrected, these variations induce fatal alignment errors during matrix operations [3]. Utilizing QGIS, all raster arrays were resampled and forcibly reprojected to the EPSG:32643 (WGS 84 / UTM Zone 43N) coordinate system.



Fig. 3. Workflow adopted for data preprocessing and machine learning implementation.

For the machine learning models executed in Google Earth Engine, continuous variables required scaling [12]. Because elevation values scale in the hundreds of meters while SPI values are fractions, distance-based algorithms would unfairly weight the larger numbers. Z-score normalization was applied to constrain the variance. The full dataset was subsequently randomized and partitioned, holding 70% for algorithmic training and securing 30% strictly for unseen validation testing to ensure adherence to the principles of empirical risk minimization.

B. System Data Flow Architecture

To systematically map the transformation of raw geospatial datasets into actionable predictive intelligence, the system architecture is defined using three levels of Data Flow Diagrams (DFD). These diagrams illustrate the logical progression of data through the predictive pipeline.

1) *DFD Level 0: Context Diagram*: The Level 0 Context Diagram (Fig. 4) establishes the highest level of abstraction for the project. The core "Flood Susceptibility Prediction System" functions as the centralized processing hub. It establishes an interface with two primary external entities. On the input side, "Geospatial Data Sources" provide the raw environmental data, including Digital Elevation Models (DEM), satellite imagery, meteorological datasets, and historical field reports [1], [2], [7]. After internal processing, the system provides "Municipal Authorities and Urban Planners" with validated susceptibility maps, ROC-AUC performance metrics, and feature importance data to support disaster mitigation strategies [4], [11].



Fig. 4. DFD Level 0 (Context Diagram) illustrating the boundary of the flood susceptibility prediction system.

2) *DFD Level 1: System Pipeline*: The Level 1 DFD (Fig. 5) decomposes the central process into a sequential Prediction Pipeline. Data standardization acts as the initial entry point, ingesting raw inputs to ensure spatial consistency across the heterogeneous datasets [3], [8]. This is followed by inventory synthesis, where historical coordinates are converted into structured training targets. The processed data is then routed into the algorithmic training phase, which serves as the computational core where models capture non-linear relationships [4], [12]. Finally, the map generation process transforms the model weights into the final geographical risk classifications distributed to end-users [13].



Fig. 5. DFD Level 1 detailing the internal four-stage prediction pipeline.

3) *DFD Level 2: Decomposition of Algorithmic Training*: To provide granular visibility into the predictive engine, the Level 2 DFD (Fig. 6) breaks down Process 3.0 (Algorithmic Training). Initially, Process 3.1 strictly partitions the preprocessed spatial data into training and validation matrices, which are stored in dedicated data repositories (D1 and D2) to ensure unbiased evaluation [1], [4]. The training matrix is then executed through three parallel modeling sub-processes: FRM execution (3.2), SVM optimization (3.3), and Random Forest ensemble generation (3.4) [9], [12]. The resulting model weights (D3) and probability arrays are finally transmitted to the ROC-AUC model evaluator for standardized performance quantification [11].

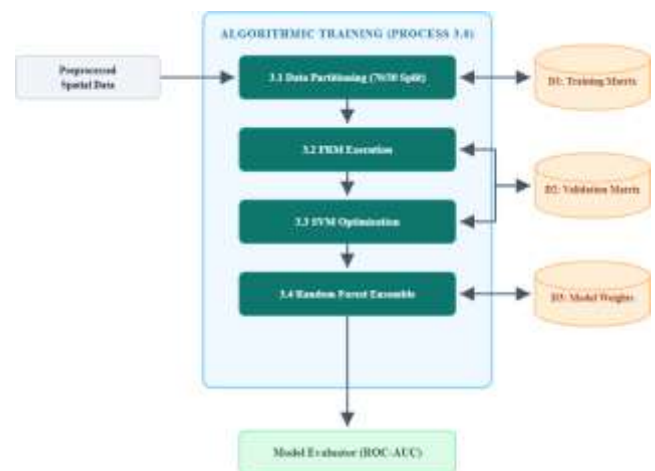


Fig. 6. DFD Level 2 providing an operational breakdown of the parallel modeling and training mechanics.

C. Frequency Ratio Model (FRM) Formulation

The FRM operates as the statistical baseline, utilizing a probabilistic framework [10]. It assumes that if empirical flood data is highly concentrated within a specific environmental subset relative to that subset's total area, the vulnerability is high. The Frequency Ratio is calculated as:

$$FR_i = \frac{(N_{f,i}/N_f)}{(N_{p,i}/N_p)} \quad (1)$$

Where $N_{f,i}$ is the count of flood occurrences in class i , N_f is the total flood occurrences, $N_{p,i}$ is the pixel count of class i , and N_p is the total pixels [14].

The spatial Flood Susceptibility Index (FSI) is then computed by linearly summing the derived ratios:

$$FSI = \sum_{i=1}^n (FR_i \times X_i) \quad (2)$$

Where X_i indicates the presence of the variable class at a given spatial coordinate.

D. Support Vector Machine (SVM) Formulation

SVM is a mathematically rigorous supervised learning model designed to construct an optimal hyper-dimensional decision boundary [13]. Given a set of training vectors x_i and their binary classifications $y_i \in \{-1, 1\}$, the standard SVM optimization problem seeks to maximize the margin separating the classes by minimizing the weight vector w :

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (3)$$

Subject to the constraint $y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i$, where ξ_i represents slack variables allowing for minor misclassifications to prevent severe overfitting, and C is the regularization penalty parameter [9].

Because environmental spatial interactions are highly non-linear, a linear hyperplane is insufficient. The Radial Basis Function (RBF) kernel was employed to project the input space into a higher dimension without explicitly calculating the transformation coordinates:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (4)$$

Where γ controls the radius of influence of individual training samples.

E. Random Forest (RF) Formulation

The Random Forest architecture leverages the mathematics of ensemble learning [12]. Implemented via the `smileRandomForest()` protocol, RF aggregates the predictions of hundreds of independently constructed decision trees.

Each tree partitions the spatial data by searching for the variable split that maximizes information gain or minimizes node impurity. The Gini impurity G for a given node with C classes is calculated as:

$$G = 1 - \sum_{i=1}^C (p_i)^2 \quad (5)$$

Where p_i is the probability of an item belonging to class i [4]. RF utilizes Bootstrap Aggregating (Bagging), meaning each tree is trained on a randomized subset of the inventory data. Furthermore, at every node split, only a random subset of the 10 conditioning factors is evaluated. This forced decorrelation ensures that highly dominant variables (like elevation)

do not mask the subtle predictive power of secondary indices (like TRI), resulting in a highly robust, low-variance predictive model.

V. EVALUATION METRICS

Model efficacy was evaluated using threshold-independent statistical metrics [3]. The confusion matrix yields True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), from which standard rates are derived:

$$TPR = \frac{TP}{TP + FN} \quad (6)$$

$$FPR = \frac{FP}{FP + TN} \quad (7)$$

These form the Receiver Operating Characteristic (ROC) curve. The Area Under the Curve (AUC) offers a definitive, aggregate measure of performance [11]. Overall Accuracy and the Cohen's Kappa Coefficient were also computed to measure precise categorization reliability while explicitly accounting for chance agreement in the spatial data.

VI. RESULTS AND DISCUSSION

The comparative evaluation yielded stark performance differences, validating the hypothesis that non-linear ensemble methods mathematically outpace traditional statistics in complex hydrological basins [4], [12].

A. Baseline Evaluation: FRM Output

The FRM achieved a baseline accuracy of 89.20%. While it successfully identified the primary inundation corridors bordering the Panchganga riverbanks, the resulting spatial output (Fig. 7) suffered from heavy visual homogenization. Because FRM calculates probabilities independently, it failed to register localized, non-linear flood pockets where minor variations in terrain ruggedness interact uniquely with dense urban land cover.

B. Support Vector Machine Results

Operating with the RBF kernel, the SVM model recorded an overall accuracy of 79.17% and an AUC of 0.81. Although the mathematical projection into high dimensional space theoretically allows for complex boundaries, the spatial mapping in Kolhapur appeared highly fragmented. SVM classification boundaries are notoriously sensitive to data scaling and geographic heterogeneity. In transitional zones where urban grid edges meet agricultural slopes the SVM optimization equation struggled to generalize, resulting in a higher false positive rate (Table II) compared to the ensemble approach.

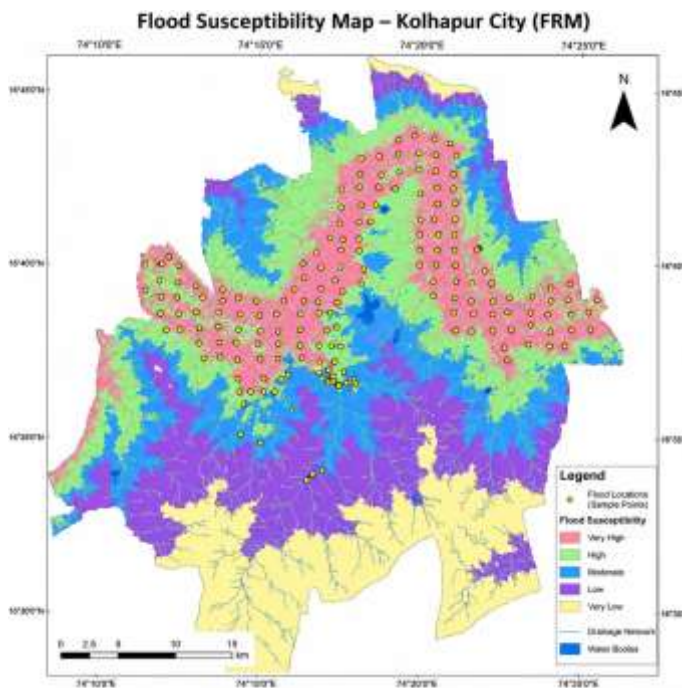


Fig. 7. Flood susceptibility map generated using the Frequency Ratio Model (FRM).

TABLE II
CONFUSION MATRIX FOR SUPPORT VECTOR MACHINE (SVM)

Actual/Predicted	Non-Flood (0)	Flood (1)
Non-Flood (0)	16	12
Flood (1)	3	41

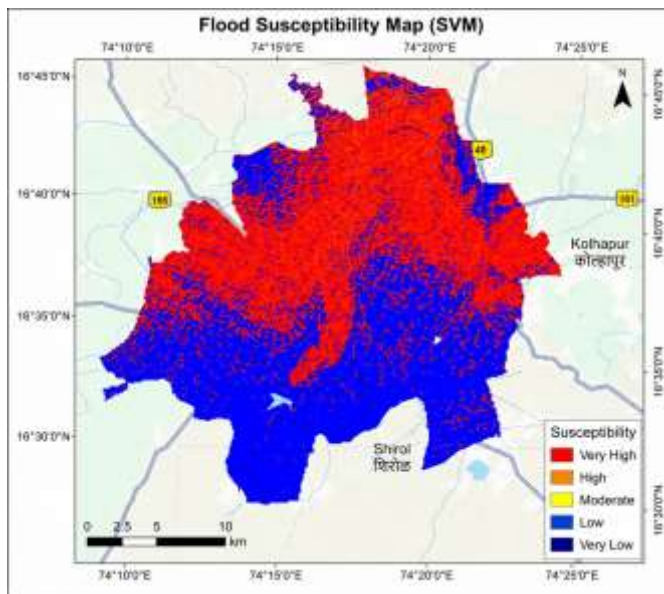


Fig. 8. Flood susceptibility map generated using the Support Vector Machine (SVM).

C. Random Forest Superiority

The Random Forest model dominated the predictive evaluation, achieving a 90.28% overall accuracy, a Kappa coefficient of 0.789, and an exceptional ROC-AUC score of 0.93. The ensemble architecture successfully bypassed the limitations of the previous models. By aggregating the outcomes of thousands of decorrelated decision trees, RF effectively smoothed the chaotic noise of the terrain data while retaining the sharp accuracy required to map specific, localized urban depressions.

TABLE III
CONFUSION MATRIX FOR RANDOM FOREST (RF)

Actual / Predicted	Non-Flood (0)	Flood (1)
Non-Flood (0)	22	6
Flood (1)	1	43

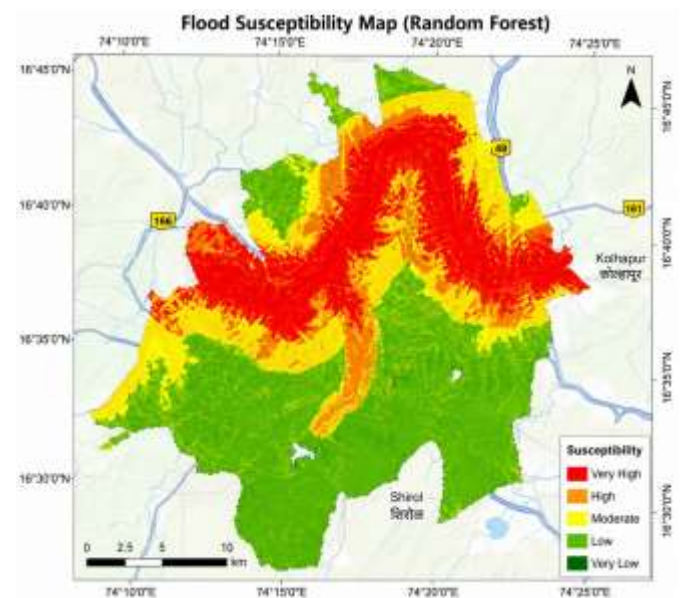


Fig. 9. Flood susceptibility map generated using the Random Forest (RF) model.

TABLE IV
PERFORMANCE EVALUATION METRICS OF IMPLEMENTED MODELS

Model	Accuracy (%)	ROC-AUC	Kappa Coefficient
FRM	89.20	0.89	-
SVM	79.17	0.81	0.534
RF	90.28	0.93	0.789

The corresponding ROC curve confirms RF as the superior diagnostic algorithm, maximizing the true positive classification while stringently controlling false positives.

D. Feature Importance Analytics

A defining advantage of tree-based ensembles is mathematical transparency [11]. The Random Forest Gini importance tracking revealed a stark hierarchy governing inundation mechanics in Kolhapur. Distance from the river constituted 48.6%

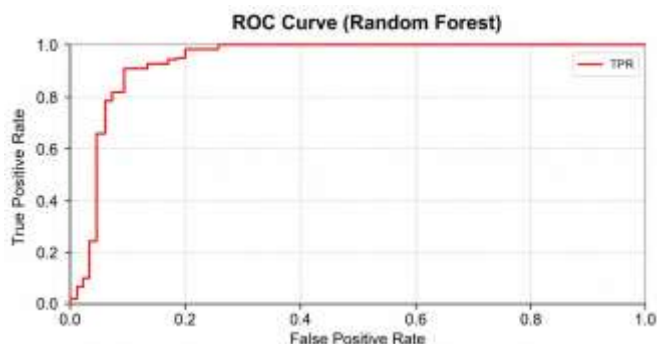


Fig. 10. ROC curve generated for the Random Forest (RF) model.

of the predictive weight, followed by Elevation at 28.7%. This empirical output verifies that despite rapid urbanization, the fundamental geomorphology of the Panchganga basin remains the primary threat vector. Secondary parameters like Terrain Ruggedness and LULC provided vital calibration data, allowing the algorithm to correctly map the moderate transition zones that standard statistics missed.

TABLE V
FEATURE IMPORTANCE OBTAINED FROM THE RF MODEL

Parameter	Contribution (%)
Distance from River	48.6
Elevation	28.7
TRI	3.1
LULC	3.1
Rainfall	2.8
Plan Curvature	2.7
Slope	1.0
TWI	0.8
SPI	0.4
Aspect	0.1

VII. CONCLUSION AND FUTURE SCOPE

Strategic mitigation of monsoon-induced inundation requires predictive models capable of processing massive, non-linear geographical matrices. This research engineered a rigorous, standardized architecture to test statistical and machine learning paradigms within Kolhapur city. By strictly controlling the preprocessing pipeline and validation constraints, the study conclusively proves that ensemble algorithmic models provide superior spatial intelligence compared to traditional methodologies.

While the Frequency Ratio Model generated a mathematically functional baseline (89.20%), it proved incapable of mapping heterogeneous local hazards. Support Vector Machine mapping (79.17%) suffered from hyper-dimensional scaling issues, producing fragmented risk zones. The Random Forest architecture, however, demonstrated exceptional predictive resilience, registering the highest accuracy (90.28%) and an ROC-AUC of 0.93. The resulting susceptibility map provides local government authorities with verified, actionable spatial

data required to regulate future urban zoning and engineer targeted drainage infrastructure.

To advance regional hazard resilience, future modeling efforts must transition from static topological susceptibility to-ward dynamic hydro-meteorological forecasting [8]. Integrating the optimized Random Forest architecture with real-time stream gauge telemetry and dynamic satellite soil-moisture indices would facilitate a live early-warning system, drastically reducing the latency of municipal disaster response protocols during peak monsoon periods.

REFERENCES

- [1] S. K. Powar and S. S. Panhalkar, "Flood-susceptibility analysis of Kolhapur City using Frequency Ratio Model," *Geomatics and Environmental Engineering*, vol. 18, no. 6, pp. 23-45, 2024.
- [2] S. S. Panhalkar and S. K. Powar, "GIS-based flood susceptibility mapping in Kolhapur region," *International Journal of Advanced Remote Sensing and GIS*, 2022.
- [3] H. Arabameri, B. Pradhan, A. Rezaei, K. Lee, and N. Saro, "GIS-based flood susceptibility mapping using frequency ratio and evidential belief function models," *Environmental Earth Sciences*, vol. 78, no. 21, 2019.
- [4] T. M. Asrade, S. A. Abebe, K. B. Tadesse, M. S. Kerebih, and T. M. Meshesha, "Flood susceptibility assessment using three machine learning techniques and comparison of their performance," *Scientific Reports*, vol. 16, no. 1, art. no. 8099, 2026.
- [5] K. Khosravi, B. T. Pham, I. Prakash, and D. T. Bui, "A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, Iran," *Science of the Total Environment*, vol. 627, pp. 744-755, 2018.
- [6] D. T. Bui, P. Tsangaratos, V. T. Nguyen, and K. Khosravi, "Comparative analysis of machine learning methods for spatial prediction of flood-prone areas," *ISPRS International Journal of Geo-Information*, vol. 9, no. 3, 2020.
- [7] A. Riche, A. Drias, M. Guermoui et al., "A novel hybrid deep-learning approach for flood-susceptibility mapping," *Remote Sensing*, vol. 16, no. 19, art. no. 3673, 2024.
- [8] O. Rahmati, H. R. Pourghasemi, and H. Zeinivand, "Flood susceptibility mapping using frequency ratio and weights-of-evidence models," *Geo-matics, Natural Hazards and Risk*, vol. 7, no. 5, pp. 1639-1661, 2016.
- [9] M. S. Tehrany, B. Pradhan, and M. N. Jebur, "Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models," *Journal of Hydrology*, vol. 512, pp. 332-343, 2014.
- [10] A. Mosavi, P. Ozturk, and K. W. Chau, "Flood prediction using machine learning models: Literature review," *Water*, vol. 10, no. 11, art. no. 1536, 2018.
- [11] A. Arabameri, B. Pradhan, K. Khosravi, and K. Lee, "Spatial modelling of flood susceptibility using machine learning models," *Journal of Hydrology*, vol. 573, pp. 928-944, 2019.
- [12] H. Chen, B. Panahi, and S. Pourghasemi, "Flood susceptibility modeling using GIS-based approaches: A case study," *Natural Hazards*, vol. 95, no. 1, pp. 123-145, 2019.
- [13] K. Khosravi, B. T. Pham, and D. T. Bui, "Hybrid machine learning approaches for flood susceptibility mapping," *Environmental Modelling and Software*, vol. 117, pp. 222-234, 2019.
- [14] S. Lee and B. Pradhan, "Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models," *Landslides*, vol. 4, no. 1, pp. 33-41, 2007.
- [15] M. B. Kia, H. Pirasteh, B. Pradhan, A. R. Mahmud, W. N. A. Sulaiman, and A. Moradi, "An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia," *Environmental Earth Sciences*, vol. 67, no. 1, pp. 251-264, 2012.