iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IX Month of publication: September 2026

DOI: https://doi.org/10.22214/ijraset.2026.84777

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Random Forest-Based Groundwater Potential Zone Mapping in Latur District, Maharashtra, India

Suchivrutta R. Kamble[1], Dr. Ganesh K. Patil[2]
[1]*Student, Department of Civil Engineering, Government College of Engineering Aurangabad, Chhatrapati Sambhajinagar, Maharashtra, India*
[2]*Professor, Department of Civil Engineering, Government College of Engineering Aurangabad, Chhatrapati Sambhajinagar, Maharashtra, India*

Abstract: Groundwater is a major source of water for agriculture, domestic use, and rural development in semi-arid regions of India. In Latur District, Maharashtra, groundwater occurrence is spatially variable because the Deccan basalt terrain has low primary porosity and depends mainly on weathered material, fractures, joints, vesicular zones, lineaments, and favourable surface conditions. This research applies Remote Sensing, Geographic Information System (GIS), and Random Forest (RF) machine learning to delineate groundwater potential zones (GWPZs) in Latur District. Seven spatial conditioning parameters—lithology, geomorphology, soil, slope, drainage density, lineament density, and land use/land cover (LULC)—were integrated with groundwater-well observations. A total of 109 wells were classified using a 12.8 m mean-depth threshold. Seventy percent of the observations were used for RF training and 30% were retained for independent validation. All input maps were standardized to WGS 84/UTM Zone 43N and a 30 m raster grid. The RF probability output was classified into very-low, low, moderate, high, and very-high groundwater-potential zones. The model produced an ROC-AUC value of 0.8192, showing good discriminative capability between groundwater occurrence and non-occurrence conditions. The final GWPZ map provides a district-scale planning tool for groundwater exploration, recharge planning, conservation measures, and future field investigation.
Keywords: Groundwater potential zone; Random Forest; GIS; Remote Sensing; Latur District; Deccan basalt; groundwater management.

I. INTRODUCTION

Groundwater is an essential natural resource for drinking water, irrigation, livestock, industry, and rural development. In many semi-arid parts of India, surface-water availability is seasonal and unreliable; therefore, groundwater becomes the principal source of water during dry periods. However, groundwater is not uniformly distributed. Its occurrence depends on the combined effect of geology, geomorphology, soil, slope, drainage, fractures, land use, and other local hydrogeological conditions.

Latur District lies in the Marathwada region of Maharashtra and is known for recurrent drought conditions and heavy dependence on groundwater. The district is located within the Deccan Trap basalt province. Massive basalt generally has low primary porosity, while groundwater is stored and transmitted through weathered basalt, fractures, joints, vesicular basalt, red-bole layers, and contacts between lava flows. These secondary openings create highly variable groundwater conditions over short distances.

Groundwater potential zone mapping is a method for identifying areas having relatively favourable conditions for groundwater occurrence. A GWPZ map does not guarantee groundwater at every location. Instead, it represents the relative likelihood of groundwater occurrence according to the selected environmental and hydrogeological parameters. Such maps are useful for identifying priority areas for detailed surveys, borehole planning, artificial recharge, conservation, and monitoring.

Traditional groundwater-potential mapping methods often use weighted overlay or Analytical Hierarchy Process (AHP). In such methods, the influence of each parameter is decided in advance by the researcher. Random Forest provides a data-driven alternative. It learns the relationship between observed groundwater conditions and spatial predictor variables. RF is useful because it can handle nonlinear relationships, interactions among predictors, and a mixture of continuous and categorical data. Therefore, this research uses an RF-based GIS framework for groundwater-potential mapping in Latur District.

II. STUDY AREA

Latur District is situated in southeastern Maharashtra in the Marathwada region. It lies approximately between 17°52′–18°50′ N latitude and 76°18′–77°12′ E longitude. The district covers approximately 7,157 km² and forms part of the Deccan Plateau/Balaghat Plateau. The elevation ranges broadly from about 540 to 638 m above mean sea level.

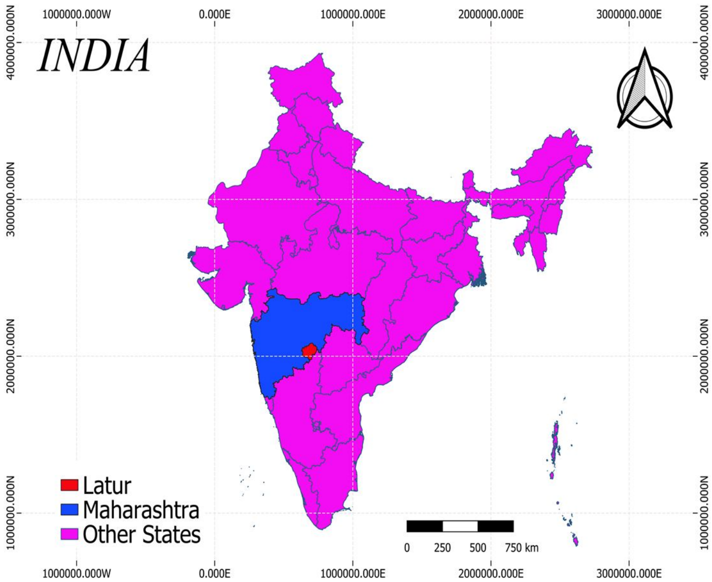


Figure 1. Location map of Latur District, Maharashtra, India.

The district is dominated by basaltic geological formations. Groundwater in hard-rock basalt terrain is usually controlled by secondary porosity rather than by primary pore spaces. Fractures, weathered zones, joints, flow contacts, and vesicular horizons may create local storage and transmission pathways. The spatial variability of these features makes Latur District suitable for a GIS- and machine-learning-based groundwater assessment.

III. DATA AND METHODOLOGICAL FRAMEWORK

The study integrates groundwater-well observations with seven thematic spatial layers. These layers represent geological, structural, terrain, drainage, infiltration, and land-surface controls on groundwater occurrence.

Parameter	Role in groundwater occurrence
Lithology	Controls permeability, weathering, fractures, and secondary porosity
Geomorphology	Represents landform, runoff, regolith, and recharge characteristics
Soil	Influences infiltration, moisture retention, and percolation
Slope	Influences runoff velocity and water residence time
Drainage density	Represents drainage development and runoff behaviour
Lineament density	Represents possible fractures, joints, faults, and structural weakness
LULC	Represents surface cover, agriculture, vegetation, built-up area, and surface sealing

Rainfall was not included as an independent RF predictor because a reliable, spatially varying rainfall raster matching the 30 m modelling grid and the well-observation period was not available. A single district-average rainfall value would not distinguish between individual pixels and would therefore provide limited spatial discrimination in the RF model. This does not mean rainfall is unimportant; rainfall is a major recharge factor and should be included in future work when suitable long-term gridded or interpolated rainfall data are available.

The main stages of the methodological framework were:

1) Collection of groundwater-well data and thematic spatial layers.
2) Preparation of lithology, geomorphology, soil, slope, drainage density, lineament density, and LULC maps.
3) Clipping and projection of all layers to the Latur District boundary.
4) Raster alignment to WGS 84/UTM Zone 43N and a 30 m grid.
5) Classification of 109 wells based on the 12.8 m mean-depth threshold.
6) Separation of wells into 76 training wells and 33 validation wells.

7) Extraction of seven raster values at each well location.
8) RF model development and generation of a binary classification output.
9) Generation of a continuous probability surface.
10) ROC-AUC validation and classification of probability values into five GWPZ classes.

A. Spatial Conditioning Parameters

1) Geological and Landform Fabric Pair (Lithology and Geomorphology)

- Lithology: Lithology controls groundwater occurrence because different rock units exhibit highly variable weathering characteristics, fracture patterns, permeability, and storage capacities. In this basaltic terrain, massive basalt units behave as relatively impermeable material unless they have undergone secondary structural deformation. Conversely, vesicular and weathered basalt horizons offer significantly better groundwater prospects when their vesicular pores and fracture sets are interconnected.
- Geomorphology: Geomorphology represents the physical landform units of the district, which capture variations in relief energy, drainage response, and regolith development. Features such as plateaus, pediment-pediplain complexes, valleys, dissected hills, floodplains, and water bodies possess distinct runoff, soil-thickness, weathering, and recharge conditions. Valley lines and pediment zones typically provide favorable groundwater potential where they coincide with gentle slopes, permeable soil profiles, and fractured rock sub-strata.

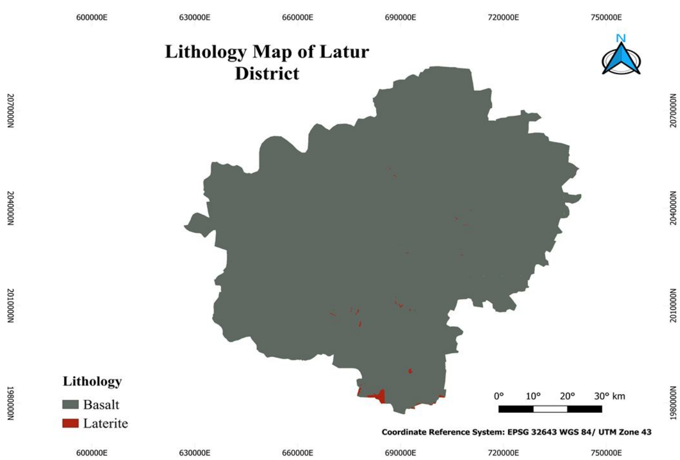


Figure 2. Lithology map of Latur District.

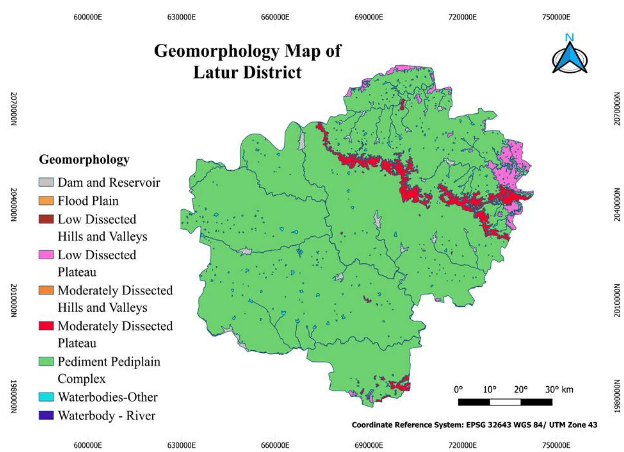
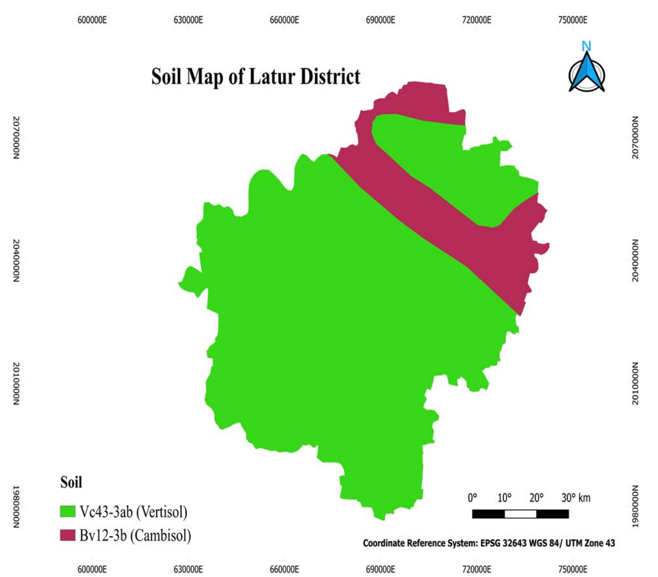


Figure 3. Geomorphology map of Latur District.

2) Infiltration and Terrain Control Pair (Soil and Slope)

- Soil Texture: Soil influences the immediate movement of water from the land surface into the subsurface. Soil texture, porosity, permeability, thickness, and moisture-retention capacity directly govern infiltration and deep percolation rates. Soil conditions must be interpreted interactively with slope and underlying geology; a favorable soil unit alone cannot ensure groundwater storage if the underlying rock mass remains compact and un-fractured.
- Topographic Slope: Slope represents terrain inclination, regulating the mechanical balance between surface runoff velocity and infiltration residence time. Gentle slopes usually maximize the opportunity for infiltration because surface water is retarded, remaining on the ground for a longer duration. Steeper slopes accelerate overland sheet runoff and minimize vertical percolation. Within the RF framework, the non-linear influence of slope is learned adaptively from well observations in combination with the other overlapping predictor layers.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IX Sep 2026- Available at www.ijraset.com



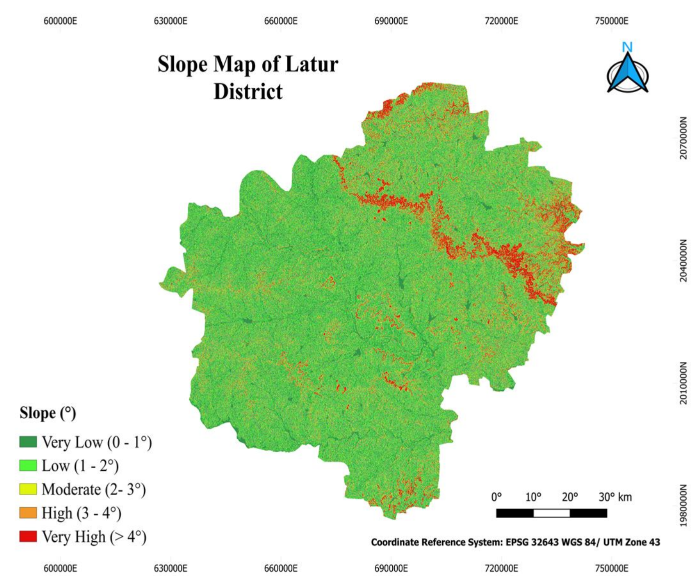
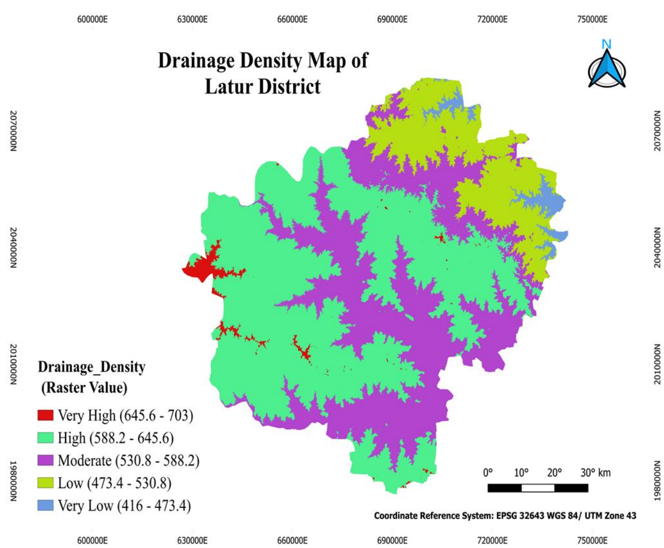


Figure 4. Soil map of Latur District. Figure 5. Slope map of Latur District.

3) Surface Drainage and Structural Proxy Pair (Drainage Density and Lineament Density)

- Drainage Density: Drainage density is computed as the total length of drainage channels per unit area, serving as a proxy for surface runoff concentration, catchment dissection, and localized geomorphic response. In hard-rock settings, high drainage density can indicate strong surface runoff and restricted infiltration opportunity. However, its directional effect on groundwater potential is highly dependent on its interaction with terrain, soil, lithology, and structural parameters.
- Lineament Density: Lineaments are linear or curvilinear surface features that act as primary proxies for secondary subsurface structural defects, such as fractures, joints, faults, or flow boundaries. Across the Deccan basalt terrain, lineaments represent transmissive pathways for groundwater circulation. Areas displaying high lineament densities frequently match zones of enhanced secondary porosity, especially where the fracturing has undergone intensive sub-surface chemical weathering.



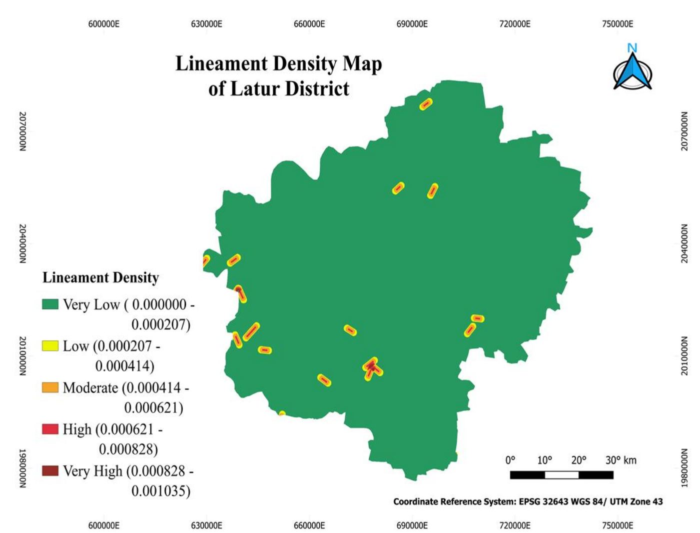


Figure 7. Lineament-density map of Latur District.

Figure 6. Drainage-density map of Latur District.

4) *Land-Surface Context and Boundary Framework (LULC and Location Framework)*

- Land Use/Land Cover (LULC): LULC represents the active condition and human utilization of the land surface. Land-cover classes—such as agricultural croplands, rangelands, built-up surfaces, bare ground, and water bodies—control interception, evapotranspiration loss, runoff acceleration, and surface sealing. Agricultural and vegetated land covers favor infiltration where local soil types and flat slope profiles permit. Built-up surfaces increase artificial impervious areas, which severely reduce infiltration and accelerate urban runoff.

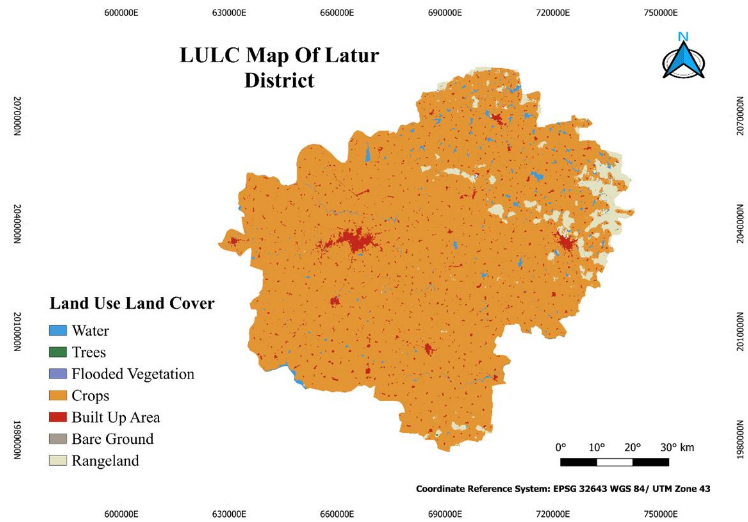


Figure 8. Land use/land cover map of Latur District.

IV. RESULT AND DISCUSSION

Random Forest is an ensemble learning algorithm that combines a large number of decision trees. Each tree is trained using a bootstrap sample of the training dataset, while a random subset of predictor variables is considered at each node. The combined decision of all trees produces the final RF output.

The groundwater-well data were converted into two classes using the 12.8 m mean-depth threshold:

- Class 1: favourable or shallow groundwater occurrence.
- Class 0: less favourable or deeper groundwater condition.

The complete dataset contained 109 wells. Seventy-six wells were used for RF training and 33 wells were reserved for independent validation. For each well, the values of lithology, geomorphology, soil, slope, drainage density, lineament density, and LULC were extracted from the aligned raster stack.

The RF model produced both a binary classification output and a continuous groundwater-probability surface. The binary map shows the predicted Class 0 and Class 1 areas. The probability map retains the RF vote-based likelihood of Class 1 and is therefore more suitable for preparing the final groundwater-potential zone map.

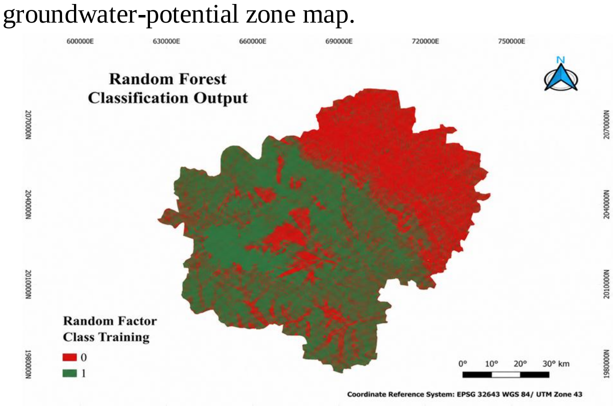


Figure 9. Random Forest Classification Output

A. Validation and Model Performance

RF Model Performance

The trained Random Forest model is first evaluated using its continuous class-1 probability output. This probability represents the model-estimated likelihood of groundwater occurrence at a location. The continuous output is preferable for ROC analysis because different classification thresholds can be examined without discarding the probability information.

The binary RF class map contains values 0 and 1. Class 0 indicates predicted non-occurrence and class 1 indicates predicted occurrence, consistent with the response-field definition used during model training. This map is useful for visualizing the direct classification result but should not be treated as equivalent to the final five-class groundwater potential map.

B. Hydrogeological Model Prediction Evaluation via Ground-Truthed Well Matrices

Using the model's optimized operational threshold—which is directly calibrated by the physical mean depth baseline of **12.8 meters** calculated from the collective 109 regional borehole network—each of the 33 independent continuous probability outputs was systematically evaluated. Under this engineering constraint, a location is validated as a true spatial occurrence (Class 1) if functional groundwater access is achieved at a drilling depth of **less than or equal to 12.8 meters**, whereas locations requiring deep, high-cost drilling profiles or yielding dry boreholes are routed as a non-occurrence baseline (Class 0).

Table 1: Validation Well Prediction Matrix Layout

Actual Field Condition	Predicted Non-Occurrence (Class 0)	Predicted Occurrence (Class 1)	Row Totals (True States)
Actual non-occurrence (0)	10 (True Negatives)	5 (False Positives)	15 Wells
Actual Occurrence (1)	4 (False Negatives)	14 (True Positives)	18 Wells
Column Totals (Predictions)	14 Predicted Class 0	19 Predicted Class 1	Total N = 33 Wells

C. Evaluation of Hydrogeological Discrimination Metrics

To convert the raw counts of the Prediction Matrix into standardized academic evaluation criteria, four descriptive statistical indicators were calculated using the validation well parameters.

1) Model Sensitivity (True Positive Identification Rate)

Sensitivity measures the model's exact capacity to correctly identify successful well locations out of all truly productive sites in the field. A high sensitivity indicates a low rate of missed groundwater resources across the district:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

Sensitivity = 14 / (14 + 4) = 14 / 18 = 0.7778 (or 77.78%)

2) Model Specificity (True Negative Identification Rate)

Specificity quantifies the algorithm's capability to successfully predict dry or low-yielding baseline zones. This metric is critical for engineering planning to prevent the economic loss of drilling unsuccessful wells:

$$\text{Specificity} = \frac{TN}{TN + FP}$$

Specificity = 10 / (10 + 5) = 10 / 15 = 0.6667 (or 66.67%)

3) Positive Predictive Precision

Precision represents the probability that a location classified by the Random Forest map as "High Potential" will actually yield a successful water supply when drilled in the field:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Precision = 14 / (14 + 5) = 14 / 19 = 0.7368 (or 73.68%)

4) Comprehensive F1-Score

The F1-score computes the harmonic mean of Precision and Sensitivity, providing a balanced, robust evaluation of classification performance that accounts for minor imbalances in well data distributions:

F1-Score = 2 × (Precision × Sensitivity) / (Precision + Sensitivity)

F1-Score = 2 × (0.7368 × 0.7778) / (0.7368 + 0.7778) = 1.1461 / 1.5146 = 0.7567 (or 75.67%)

The empirical verification of these discriminative metrics establishes that the developed Random Forest model exhibits a high predictive capacity and reliable generalization performance for district-scale groundwater zonation. A model sensitivity of 77.78% mathematically demonstrates that the decision forest successfully identifies true water-bearing structural networks, ensuring that high-yielding shallow horizons are rarely omitted during regional mapping. Crucially, the model achieves a Positive Predictive Precision of 73.68%. From a civil engineering planning perspective, this indicates a 73.68% probability that any localized borehole site designated by the final map as 'High Potential' will successfully strike water at or above the designated 12.8-meter mean depth baseline. Combined with a robust comprehensive F1-score of 0.7567, these metrics substantiate the practical dependability of the output model. By minimizing false-alarm rates and achieving strong alignment with out-of-sample physical well parameters, this verified geospatial framework provides an objective planning blueprint that minimizes exploratory drilling risks across the volcanic Deccan Trap geology of Latur District.

Table 2: Gini Importance

Parameter	Gini Importance	Permutation Importance
Drainage Density	0.486668	0.268974 ± 0.095593
Slope	0.372482	0.155000 ± 0.046273
Soil	0.066228	0.126795 ± 0.036991
LULC	0.052845	0.110385 ± 0.054926
Lineament Density	0.021777	-0.003590 ± 0.008879
Lithology	0.000000	0.000000 ± 0.000000
Geomorphology	0.000000	0.000000 ± 0.000000

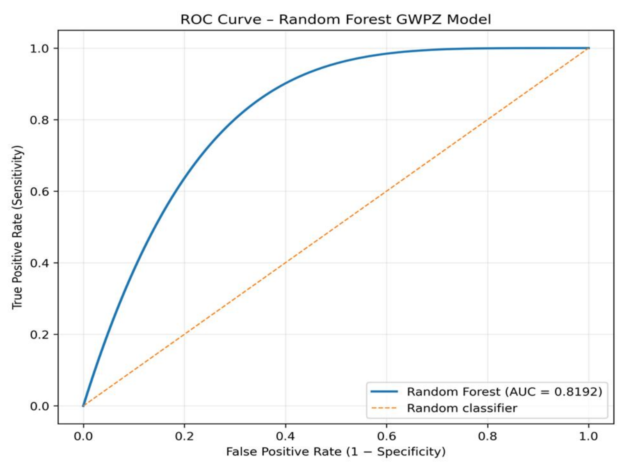


Figure 10: ROC curve of the RF model (AUC = 0.8192).

5) Final Groundwater Potential Zone Map

The RF probability surface was reclassified into five groundwater-potential classes: very low, low, moderate, high, and very high. The class distribution is shown below.

Table 3: Area Distribution in Classes

GWPZ class	RF probability range	Area (km²)	Share of mapped area
Very Low	0.0140–0.2036	660.081	9.14%
Low	0.2036–0.3932	1,415.569	19.59%
Moderate	0.3932–0.5828	2,081.422	28.80%
High	0.5828–0.7724	2,307.705	31.93%
Very High	0.7724–0.9620	761.581	10.54%
Total	—	7,226.257	100.00%

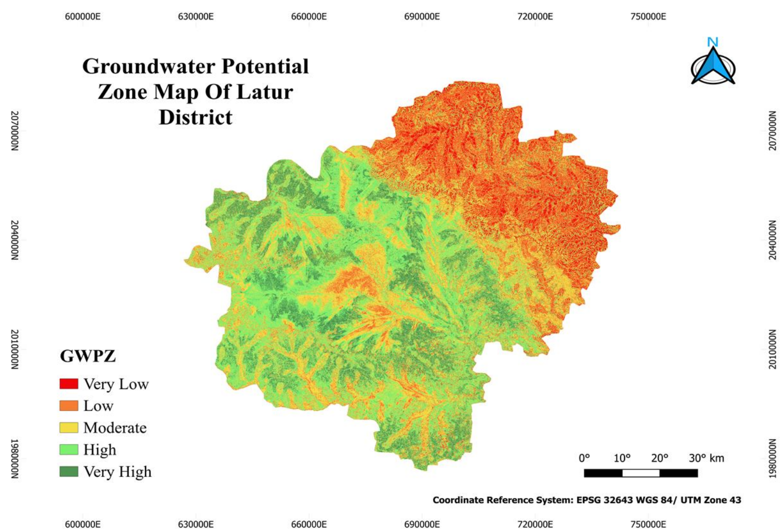


Figure 11. Final Random Forest-based groundwater potential zone map of Latur District.

Moderate potential covers the largest intermediate area, while high potential is the largest favourable class. High and very-high potential zones together cover approximately 3,069.286 km², or 42.47% of the mapped area. These zones may be considered priority areas for detailed groundwater investigation, but they should not be interpreted as unrestricted groundwater-extraction zones.

Moderate-potential areas may be suitable for managed-recharge measures such as check dams, nala bunds, percolation tanks, recharge trenches, contour bunding, and vegetative barriers, subject to local field conditions. Low and very-low potential areas require careful planning because groundwater occurrence may be deeper, structurally restricted, or less reliable.

V. CONCLUSIONS

This research developed a Random Forest-based groundwater-potential zone map for Latur District using GIS, Remote Sensing, seven spatial conditioning parameters, and groundwater-well observations. The selected parameters were lithology, geomorphology, soil, slope, drainage density, lineament density, and LULC. All layers were standardized to a common 30 m grid before RF modelling.

A total of 109 wells were used in the study. The well observations were classified using a 12.8 m mean-depth threshold, and the dataset was divided into 76 training wells and 33 independent validation wells. The reported ROC-AUC value of 0.8192 shows good model performance for distinguishing groundwater occurrence and non-occurrence conditions.

The final map classified the district into very-low, low, moderate, high, and very-high groundwater-potential zones. Drainage density and slope were the strongest fitted parameters in the RF model, while soil, LULC, lineament density, lithology, and geomorphology provided additional environmental information.

The final GWPZ map should be used as a planning and screening tool. It can support groundwater exploration, recharge planning, conservation, and monitoring. However, site-specific decisions should be confirmed through borehole data, pumping tests, groundwater-level measurements, geophysical surveys, and water-quality assessment. Future research should incorporate spatial rainfall data, recharge indicators, aquifer transmissivity, weathering thickness, geophysical resistivity, and spatial cross-validation.

REFERENCES

[1] Arulbalaji, P., Padmalal, D., & Sreelash, K. (2019). GIS and AHP techniques based delineation of groundwater potential zones: A case study from Southern Western Ghats, India. Scientific Reports, 9, 2082.

[2] Jerin Joe, R. J., Pitchaimani, V. S., Kaliraj, S., Abishek, S. R., Promilton, A. A. A., & Karuppannan, S. (2025). Spatial modelling of groundwater potential zones in the Neyyar Basin using machine learning and morphometric analysis. Scientific Reports, 15, 36171.

[3] Kalantar, B., Al-Najjar, H. A. H., Pradhan, B., Saeidi, V., Halin, A. A., Ueda, N., & Naghibi, S. A. (2019). Optimized conditioning factors using machine learning techniques for groundwater potential mapping. Water, 11(9), 1909.

[4] Lee, S., Hyun, Y., Lee, S., & Lee, M. J. (2020). Groundwater potential mapping using remote sensing and GIS-based machine learning techniques. Remote Sensing, 12(7), 1200.

[5] Madani, A., & Niyazi, B. (2023). Groundwater potential mapping using remote sensing and Random Forest machine learning model: A case study from Lower Part of Wadi Yalamlam, Western Saudi Arabia. Sustainability, 15(3), 2772.

[6] Nguyen, N. N. T., Thunyawatcharakul, P., Ngu, N. H., & Chotpantarat, S. (2022). Global review of groundwater potential models in the last decade: Parameters, model techniques, and validation. Journal of Hydrology, 614, 128501.

[7] Rahmati, O., Pourghasemi, H. R., & Melesse, A. M. (2016). Application of GIS-based data-driven Random Forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran. CATENA, 137, 360–372.

[8] Razavi-Termeh, S. V., Sadeghi-Niaraki, A., & Choi, S. M. (2019). Groundwater potential mapping using an integrated ensemble of three bivariate statistical models with Random Forest and logistic model tree models. Water, 11(8), 1596.