

Groundwater Potential and Managed Aquifer Recharge Suitability Assessment using GIS-AHP and Machine Learning in the upstream part of Awash River, Ethiopia

Muauz Redda

muauzamare79@gmail.com

Addis Ababa University

Behailu Birhanu

Addis Ababa University

Bedru Hussien

Addis Ababa Science and Technology University

Research Article

Keywords: Analytical Hierarchy Process, Ensemble Modeling, Groundwater Potential Index, Machine Learning, Managed Aquifer Recharge (MAR), ROC–AUC, Precision–Recall Curve

Posted Date: March 11th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8850489/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Groundwater Potential and Managed Aquifer Recharge Suitability Assessment using GIS-AHP and Machine Learning in the upstream part of Awash River, Ethiopia

Muauz Amare Redda^{a,*}, Behailu Birhanu^b, Bedru Hussien Mohammed^c

a Ethiopian Institute of Water Resource, Addis Ababa University, Addis Ababa, Ethiopia, ORCID iD: <https://orcid.org/0000-0006-8220-4470>

b School of Earth Sciences, Addis Ababa University, Addis Ababa, Ethiopia, ORCID iD: <https://orcid.org/0000-0002-6934-8819>

c School of Earth Sciences and Engineering, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia

**Corresponding author: Email address: muauzamare79@gmail.com*

Abstract

Managed aquifer recharge (MAR) is a key strategy for enhancing groundwater storage, reducing overexploitation, and improving climate resilience in semiarid regions. However, its effectiveness depends on accurate identification of hydrogeologically suitable recharge zones. This study developed an integrated Geographic Information System–Analytical Hierarchy Process (GIS–AHP) and ensemble machine learning (ML) framework to assess groundwater potential and MAR suitability in the upstream part of the Awash River in Ethiopia. Eleven thematic layers were weighted using AHP to derive a Groundwater Potential Index (GPI). Geophysical parameters from vertical electrical soundings, transmissivity from pumping tests, and borehole yield data were used for the calibration and independent validation. Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost) models were trained using geophysical predictors and GPI, and combined into an ensemble MAR probability model. The AHP-derived GPI exhibited a good predictive performance against the observed borehole yield (receiver operating characteristic area under the curve, ROC–AUC = 0.731; precision–recall area under the curve, PR–AUC = 0.79). The ensemble ML model showed excellent agreement with the observed MAR conditions (ROC, AUC = 0.913; PR, AUC = 0.946). High to very high MAR suitability zones cover approximately 39% of the basin and are predominantly associated with fractured volcanic formations, moderate slopes, and high lineament density, whereas low-suitability areas correspond to steep terrain and urbanized regions. Integrating expert-based multi-criteria decision analysis with data-driven ML substantially improved MAR site identification under complex volcanic hydrogeological conditions. The proposed framework provides a robust and transferable decision support tool for basin-scale MAR planning.

Keywords: Analytical Hierarchy Process; Ensemble Modeling; Groundwater Potential Index; Machine Learning; Managed Aquifer Recharge (MAR); ROC–AUC; Precision–Recall Curve.

1. Introduction

Groundwater provides a critical buffer against hydrological variability in semi-arid environments, where surface water availability is increasingly unreliable owing to climate variability, land-use change, and growing water demand. In such regions, sustainable groundwater management depends not only on estimating available resources but also on identifying spatially favorable zones for recharge, storage, and abstraction under complex subsurface conditions.

Across Sub-Saharan Africa, groundwater reliance has intensified in response to recurrent droughts, rapid population growth, and expanding irrigated agriculture (MacDonald, A. M., Bonsor, H. C., Ó Dochartaigh, B. E., & Taylor, 2012; Taylor, R. G., Todd, M., Kongola, L., 2013). Ethiopia exemplifies this trend, particularly within the Main Ethiopian Rift system, where volcanic aquifers supply water to major urban centers and agricultural development. These aquifers exhibit strong spatial heterogeneity caused by variations in lithology, fracture density, weathering depth, and tectonic structure, resulting in highly variable groundwater occurrence and productivity (Ayenew, T., Demlie, M., & Wohnlich, 2008; Kebede, 2013). The upper Awash River sub-basin is among Ethiopia's most socioeconomically important catchments, supporting domestic, industrial, and agricultural water demands. Despite its strategic role, groundwater development in the basin remains constrained by the limited characterization of fracture-controlled flow systems, uneven recharge distribution, and insufficient integration of geophysical, hydrogeological, and spatial decision support tools (Ayenew, T., 2009). As the water demand continues to increase, these limitations increase the risk of inefficient and unsustainable groundwater exploitation.

Spatial groundwater potential mapping has emerged as an effective approach to synthesize multiple controlling factors into decision-oriented frameworks. Multi-criteria decision analysis techniques, particularly the Analytical Hierarchy Process (AHP), enable transparent weighting of thematic layers while explicitly evaluating internal consistency, making them well suited to structurally complex volcanic terrains (Rahmati, O., Pourghasemi, H. R., & Melesse, 2015; Saaty, 1980). However, AHP-derived groundwater potential indices are often semiquantitative and require independent validation to ensure hydrogeological realism and practical applicability.

Managed Aquifer Recharge (MAR) is increasingly recognized as a viable strategy for enhancing groundwater resilience in water-stressed basins. Nevertheless, MAR performance is highly sensitive to subsurface transmissivity, storage capacity, and hydraulic connectivity, particularly in fractured volcanic aquifers where inappropriate site selection can lead to limited recharge efficiency or rapid percolation losses (Dillon, P., Stuyfzand, P., Grischek, T., Lluria, M., Pyne, R.D.G., Jain, R.C., Bear, J., 2019). Consequently, robust spatial identification of hydraulically favorable and operationally reliable MAR zones is a prerequisite for effective recharge planning.

Recent advances in machine learning have introduced powerful tools for modeling the non-linear relationships between hydrogeological variables and groundwater system responses. Ensemble learning approaches, which integrate multiple algorithms, have demonstrated improved robustness and predictive reliability in groundwater potential and recharge suitability mapping, particularly under data scarcity and class imbalance (Naghibi, S. A., Moghaddam, D. D., Kalantar, B., Pradhan, B., & Kisi, 2017a; Rahmati, O., Kornejady, A., Samadi, M., et al., 2020). When combined with physically informed indices, such approaches offer a promising pathway for improving groundwater and MAR decision making.

Despite these advances, few studies have explicitly integrated AHP-derived groundwater potential indices with ensemble machine learning models and independent yield-based validation in fractured volcanic aquifers. In the upper Awash River sub-basin, such integration is essential to distinguish zones that are not only hydraulically favorable but also practically viable for groundwater development and recharge enhancement.

Accordingly, this study aims to: (i) develop an AHP-based groundwater potential index (GPI) for the upper Awash River sun basin; (ii) integrate the GPI with key hydrogeological variables using an ensemble machine learning framework to predict MAR suitability; (iii) evaluate model performance using receiver operating characteristic (ROC), precision–recall (PR), and confusion matrix analyses; and (iv) validate spatial predictions against independent borehole yield data to ensure hydrogeological consistency and decision relevance for sustainable groundwater management.

2. Study Area

The upper Awash River Sub-basin is located in central Ethiopia between latitudes 8°07'07" North and 9°55'11" N, and longitudes 37°51'09" East and 39°39'50" E (Fig.1). The basin covers an area of approximately 11,697 km² and includes the highland, escarpment, and lowland physiographic regions. The elevation ranges from approximately 1,579 to 3557m above sea level. The area experiences a unimodal rainfall pattern, with an average annual precipitation of 1,078 mm and an average annual temperature of 18.6°C. The geology is dominated by Quaternary basalts, rhyolites, ignimbrites, and fractured basement rocks, which strongly control the groundwater occurrence (Amare et al., 2025).

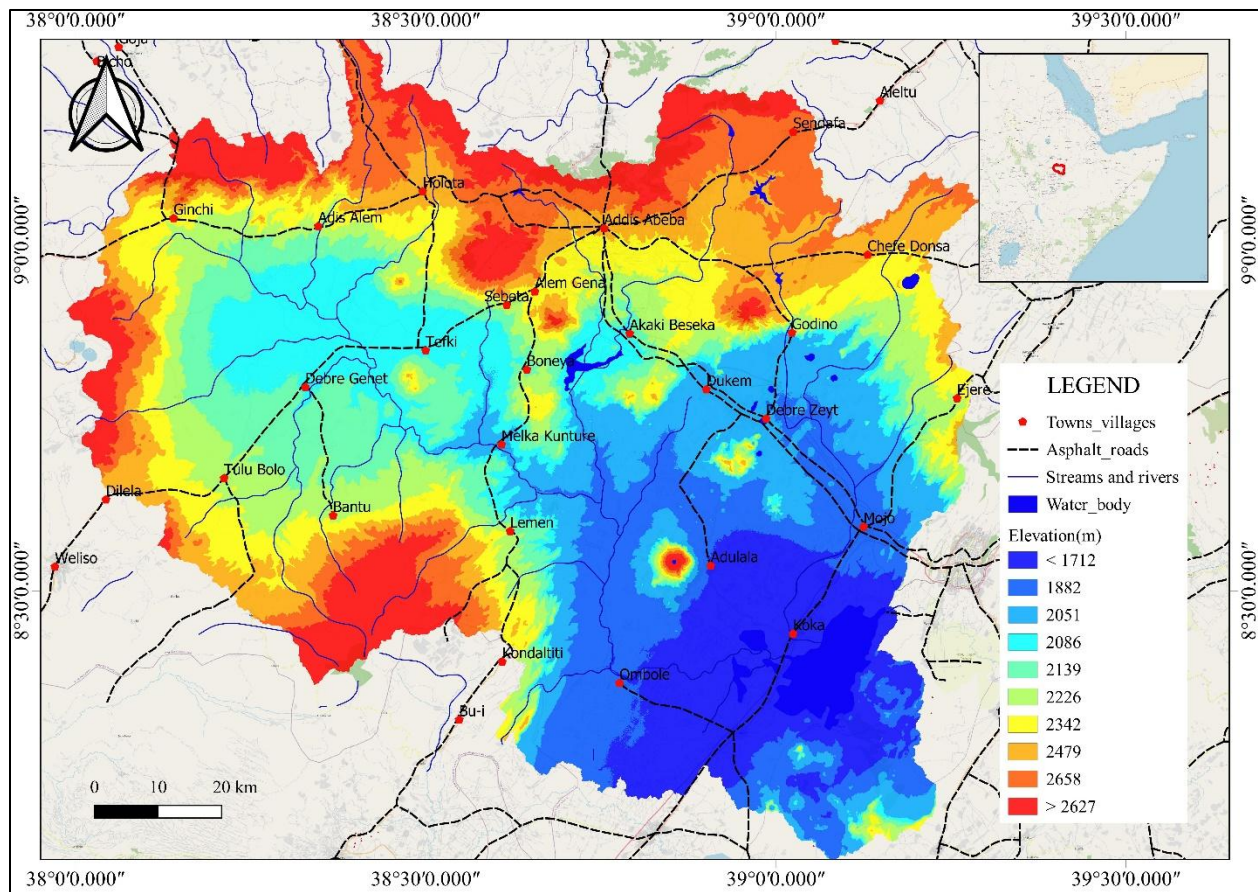


Fig.1. Location map of the upper Awash River sub-basin, showing the major rivers, administrative boundaries, and monitoring wells.

3. Data and Methodology

3.1 Data layers

The geographic location of the study area, together with the spatial distribution of the Vertical Electrical Sounding (VES) points, transmissivity measurements, and well-yield observation sites, are shown in Fig.2.

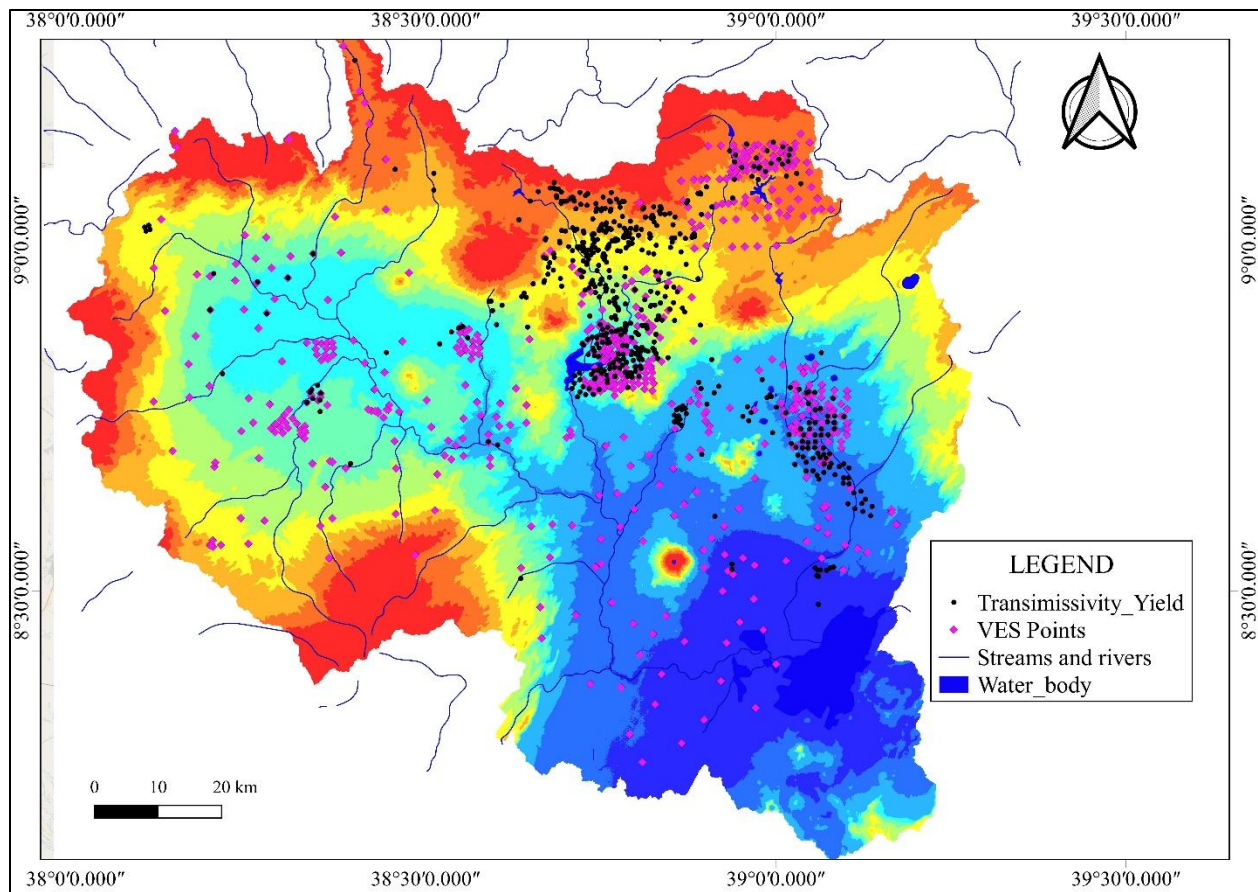


Fig.2. Map of spatial distribution of Vertical Electrical Sounding (VES) points, transmissivity measurements, and well-yield observation points

Multiple datasets were compiled from the field surveys, government agencies, and remote sensing sources (Table 1). Vertical Electrical Sounding (VES) data were used to derive subsurface resistivity and saturated thickness, pumping-test transmissivity data provided hydraulic calibration, and well-yield records were used as an independent productivity indicator, following approaches adopted in recent hydrogeophysical studies (Bernardi, L., Callegari, M., Riva, M., and Guadagnini, A., 2020; Rahmati, O., Samadi, M., Shahabi, H., Azareh, A., Rafiei-Sardooi, E., Melesse, A.M., and Pradhan, B., 2015; Yeh, H.-F., Wang, J., and Shih, H.-C., 2015). Ancillary thematic layers, including slope, drainage density, lineament density, land use/land cover, soil infiltration capacity, rainfall, and normalized difference vegetation index (NDVI), were derived from satellite imagery and national geospatial datasets.

Table 1. Thematic layers used for groundwater potential and MAR suitability assessment

Sn	Thematic Layer	Data Source	Resolution / Scale	Role in Analysis
1	Slope	SRTM Digital Elevation Model (USGS)	30 m	Controls surface runoff and infiltration; gentle slopes favor recharge and MAR suitability

2	Elevation	SRTM Digital Elevation Model (USGS)	30 m	Influences groundwater flow direction, recharge distribution, and topographic controls
3	Land Use / Land Cover (LULC)	Landsat 8 OLI (USGS)	30 m	Regula
4	Geology	Muauz et.al(2024)	1:250,000	Determines aquifer lithology, permeability, and structural controls on groundwater occurrence
5	Drainage Density	Derived from DEM	30 m	High drainage density indicates high runoff and low infiltration potential
6	Groundwater Depth (GWD)	Field well inventory data	Point-based	Indicates depth to water table; shallow groundwater enhances recharge feasibility
7	Soil Type	FAO Harmonized World Soil Database	250 m	Controls infiltration rate, storage capacity, and percolation potential
8	Normalized Difference Vegetation Index (NDVI)	Landsat 8 OLI	30 m	Proxy for soil moisture and vegetation vigor linked to recharge conditions
9	Standardized Precipitation Index (SPI)	Meteorological station data	Station-based	Represents long-term precipitation variability influencing recharge potential
10	Rainfall	National Meteorological Agency of Ethiopia	Point→ IDW	Primary source of groundwater recharge and MAR water availability
11	Topographic Wetness Index (TWI)	Derived from DEM	30 m	Indicates spatial moisture accumulation and potential recharge zones

All spatial datasets were projected onto a common coordinate reference system (UTM Zone 37N) and resampled to uniform spatial resolution prior to analysis to ensure spatial consistency (Eastman, 2016).

3.2 Geophysical data processing and aquifer parameter derivation

Vertical Electrical Sounding (VES) data were interpreted using a layered earth resistivity model to estimate aquifer resistivity (ρ) and saturated thickness (b). For each VES location, the deepest interpreted layer with measurable thickness was assumed to represent the principal aquifer horizon, consistent with hydrogeophysical investigations conducted in volcanic terrains (Barker, 1989; Ngugi et al., 2020; Mohamed et al., 2022). The derived aquifer resistivity and saturated thickness values were spatially referenced using VES coordinates and subsequently integrated into the GIS environment. These geophysical parameters constitute the primary subsurface predictors employed in both

groundwater potential assessment and managed aquifer recharge (MAR) suitability modeling (Dimitriadis et al., 2021; Rahmati et al., 2020).

3.3 Transmissivity and yield integration

The observed transmissivity values (T , m^2/day) derived from the pumping-test data were spatially integrated with the VES locations using a nearest-neighbor approach. For each VES point, the closest transmissivity measurement within a 5-km search radius was assigned, a threshold commonly used in regional aquifer characterization (Sener et al., 2018; Adiat et al., 2012). Well-yield observations were spatially joined to VES locations using the same criteria. VES points lacking nearby transmissivity or yield information were excluded from subsequent machine-learning (ML) modeling to avoid the propagation of uncertainty (Sun et al., 2021).

3.4 Analytical Hierarchy Process (AHP) and consistency evaluation

A multi-criteria Analytical Hierarchy Process (AHP) was applied to assign weights to 11 groundwater recharge influencing factors, including slope, land use/land cover (LULC), geology, drainage density, groundwater depth, elevation, NDVI, soil type, standardized precipitation index (SPI), topographic wetness index (TWI), and rainfall (Saaty, 1980; Malczewski, 2006; Rahmati et al., 2016) Table 2).

Table 2. Criteria weights used in the Analytical Hierarchy Process (AHP) for groundwater potential and MAR suitability assessment.

Criterion	Weight
Slope	0.18
LULC	0.12
Geology	0.10
Drainage Density	0.08
Groundwater Depth	0.13
Elevation	0.07
NDVI	0.05
Soil	0.06
SPI	0.07
TWI	0.06
Rainfall	0.08

The reliability of expert judgments was tested using the Consistency Ratio (CR). A pairwise comparison matrix was constructed, and λ_{max} , Consistency Index (CI), and Consistency Ratio (CR) were computed using the standard AHP formulas. A $CR \leq 0.1$ indicates acceptable consistency (Saaty & Vargas, 2012). In this study, the CR was < 0.1 , confirming the reliability of the derived weights for the weighted-overlay analysis.

To ensure the reliability of expert judgment, CI and CR were calculated as follows:

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad 1$$

$$CR = \frac{CI}{RI} \quad 2$$

3.5 Groundwater Potential Index (GPI) Mapping using AHP

The groundwater potential was evaluated using the MCDA based on AHP. Thematic layers representing slope, lithology, lineament density, drainage density, soil infiltration capacity, rainfall, and LULC were standardized and weighted using pairwise comparison matrices (Rahmati et al., 2015; Arabameri et al., 2020; Chenini & Ben Mammou, 2010). The weighted layers were integrated using a linear weighted overlay to compute the Groundwater Potential Index (GPI) using Equation 3:

$$GPI = \sum_{i=1}^n \omega_i \times r_i \quad 3$$

where ω_i represents the normalized weight of criterion i and r_i denotes its standardized rank.

The resulting GPI raster was classified into Very Low to Very High potential classes by using natural breaks.

3.6 Validation using well yield data

The predicted groundwater potential was validated against the observed well yields from boreholes distributed across the study area following standard practice (Rahmati et al., 2020; Mogaji et al., 2021). Each well location was spatially extracted from the GPI map and assigned to a predicted class. The observed yields were categorized according to Table 3, and the agreement was summarized in a confusion matrix. Additional validation involved correlation analysis between the continuous GPI and yield.

Table 3. Yield-based classification for MAR suitability and groundwater productivity.

Class	Yield (L/s)	Hydrogeological Meaning
Very Low	< 5	Weak aquifers, clays
Low	5–12	Moderately weathered basalt
Moderate	12–31.2	Fractured volcanic
High	31.2–60	Strongly fractured basalt
Very High	> 60	Deep fracture networks

3.7 Definition of MAR suitability observation variable

A binary MAR suitability index (MAR_obs) was defined using the transmissivity and yield thresholds. Locations with $T \geq T_{0.6}$ and $Y \geq Y_{0.6}$ were classified as suitable (1) or otherwise unsuitable (0). Percentile-based criteria reduce subjectivity and mitigate class imbalances (Shahraki et al., 2021; Rahmati et al., 2020).

This percentile-based definition minimizes subjectivity, maintains physical relevance, and avoids the class imbalance issues commonly encountered in MAR suitability studies.

3.8 Machine Learning Model Development

3.8.1 Predictor selection

The selection of predictor variables was guided by hydrogeological relevance, data availability, and statistical independence from the response variable to ensure a physically interpretable and robust machine-learning framework. The final predictor set comprised aquifer resistivity (ρ), saturated thickness (b), observed transmissivity (T), and the Groundwater Potential Index (GPI). Aquifer resistivity and saturated thickness represent subsurface lithological properties and groundwater storage capacity, respectively, and are widely used hydrogeophysical indicators of aquifer potential in volcanic and fractured terrains (Barker, 1989; Mohamed, A.M., Elmahdy, S.I., and Mohamed, A.E., 2022; Ngugi, P.W., Mutua, B.M., and Kiema, J.B.K., 2020). Transmissivity directly characterizes aquifer hydraulic productivity and governs the capacity of the subsurface to accept and transmit recharge water, making it a key parameter for MAR suitability assessments (Rahmati, O., Kornejady, A., Samadi, M., et al., 2020; Şener, E., Şener, Ş., Davraz, A., and Varol, S., 2018). The GPI integrates multiple surface and subsurface recharge-controlling factors derived through multi-criteria decision analysis, and provides a spatially continuous proxy for recharge favorability (Rahmati, O., Kornejady, A., Samadi, M., Nobre, A.D., Melesse, A.M., 2016). Together, these predictors capture complementary hydraulic, structural, and recharge-related controls on groundwater availability, rendering them well suited for MAR suitability modeling in hydrogeologically heterogeneous volcanic aquifer systems.

3.8.2 Model training and validation strategy

The selection of predictor variables was guided by hydrogeological relevance, data availability, and statistical independence from the response variable to ensure a physically interpretable and robust machine-learning framework. The final predictor set comprised aquifer resistivity (ρ), saturated thickness (b), observed transmissivity (T), and the Groundwater Potential Index (GPI). Aquifer resistivity and saturated thickness represent subsurface lithological properties and groundwater storage capacity, respectively, and are widely used hydrogeophysical indicators of aquifer potential in volcanic and fractured terrains (Barker, 1989; Mohamed, A.M., Elmahdy, S.I., and Mohamed, A.E., 2022; Ngugi, P.W., Mutua, B.M., and Kiema, J.B.K., 2020). Transmissivity directly characterizes aquifer hydraulic productivity and governs the capacity of the subsurface to accept and transmit recharge water, making it a key parameter for MAR suitability assessments (Rahmati, O., Kornejady, A., Samadi, M., et al., 2020; Şener, E., Şener, Ş., Davraz, A., and Varol, S., 2018). The GPI integrates multiple surface and subsurface recharge-controlling factors

derived through multi-criteria decision analysis, and provides a spatially continuous proxy for recharge favorability (Arabameri, A., Rezaei, A., Cerdà, A., Conoscenti, C., Van Oost, 2020; Rahmati, O., Kornejady, A., Samadi, M., Nobre, A.D., Melesse, A.M., 2016). Together, these predictors capture complementary hydraulic, structural, and recharge-related controls on groundwater availability, rendering them well suited for MAR suitability modeling in hydrogeologically heterogeneous volcanic aquifer systems.

3.8.3 Ensemble modeling

The ensemble MAR probability was obtained by averaging the Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost) probabilities (Dietterich, 2000; Sun, A.Y., Scanlon, B.R., Zhang, Z., Walling, D., and Mohanty, B.P., 2021). This ensemble approach leverages the complementary strengths of tree-based, kernel-based, and boosting algorithms, leading to an improved generalization performance, particularly in data-limited and hydrogeologically heterogeneous environments.

3.9 Model performance evaluation

The performance was evaluated using ROC-AUC, PR-AUC, Overall Accuracy (OA), producer's accuracy (PA), user's accuracy (UA), F1-score, and κ , which are commonly used in groundwater ML validation (Naghibi, S. A., Moghaddam, D. D., Kalantar, B., Pradhan, B., & Kisi, 2017b; Rahmati, O., Kornejady, A., Samadi, M., et al., 2020).

(1) Overall Accuracy (OA)

Overall Accuracy (OA) represents the proportion of correctly classified samples relative to the total number of validation samples and provides a global measure of model performance. The OA was calculated using Equation (6):

$$OA = \frac{\sum_{i=1}^r n_{ii}}{N} \quad 6$$

Where n_{ii} denotes the number of correctly classified samples in class i , i is the number of classes, and N is the total number of validation samples. OA reflects the overall correctness of the model but does not account for class imbalance. The interpretation of OA values used in this study is summarized in Table 4.

Table 4. Proportion of correctly classified samples across all classes.

OA Range	Interpretation
> 0.90 (90–100%)	Excellent classification
0.80–0.90	Very good
0.70–0.80	Good

OA Range	Interpretation
0.60–0.70	Moderate
0.50–0.60	Fair
< 0.50	Poor

(2) Producer’s Accuracy (PA)

The producer’s accuracy (PA), also referred to as recall or sensitivity, evaluates how effectively the model identifies the reference samples of a given class. The PA was computed using Equation 7:

$$PA = \frac{n_{ii}}{n_{i+}} \quad 7$$

Where n_{i+} is the total number of observed samples belonging to class i . PA quantifies omission error, indicating the proportion of true samples missed by the model. The classification performance levels associated with PA are provided in Table 5.

Table 5. Interpretation of Producer’s Accuracy (PA)

PA Range	Interpretation
> 0.90	Very high detection
0.70–0.90	High
0.50–0.70	Moderate
0.30–0.50	Low
< 0.30	Very low / weak detection

(3) User’s Accuracy (UA)

User accuracy (UA), also known as precision, measures the reliability of the predicted classes by indicating the proportion of correctly predicted samples within a class. The UA is expressed as Equation 8:

$$UA_i = \frac{n_{ii}}{n_{+i}} \quad 8$$

Where n_{+i} represents the total number of samples predicted as class i . UA quantifies commission error and reflects the confidence users can place in predicted MAR suitability classes. The interpretation of UA values is presented in Table 6.

Table 6. Interpretation of User’s Accuracy

UA Range	Interpretation
> 0.90	Very high reliability
0.70–0.90	High
0.50–0.70	Moderate
0.30–0.50	Low
< 0.30	Very low reliability

(4) F1-Score

The F1-score integrates the producer and user accuracies into a single metric by calculating their harmonic mean, as shown in Equation 9:

$$F_{1i} = 2 \times \frac{UA_i \times PA_i}{UA_i + PA_i} \quad 9$$

The F1-score is particularly informative under imbalanced class distributions, which are common in groundwater and MAR suitability mapping. The performance categories for the F1-score are provided in Table 7.

Table 7. Interpretation of F1-score

F1-Score Range	Interpretation
> 0.90	Excellent class performance
0.70–0.90	Strong
0.50–0.70	Moderate
0.30–0.50	Weak
< 0.30	Very weak performance

(5) Kappa Coefficient (κ)

Cohen's kappa coefficient (κ) evaluates the agreement between observed and predicted classifications, while correcting for agreement expected by random chance. The kappa coefficient was calculated using Eq. (10).

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad 10$$

- where the observed agreement (P_o) is defined as Equation 11:

$$P_o = \frac{\sum_{i=1}^r n_{ii}}{N} \quad 11$$

and the expected agreement by chance (P_e) is calculated as Equation 12:

$$P_e = \sum_{i=1}^n \left(\frac{n_{i+}}{N} \times \frac{n_{+i}}{N} \right) \quad 12$$

Thus, the complete Kappa expression becomes Equation 13:

$$\kappa = \frac{\frac{\sum_{i=1}^i n_{ii}}{N} - \sum_{i=1}^i \left(\frac{n_{i+}}{N} \cdot \frac{n_{+i}}{N} \right)}{1 - \sum_{i=1}^i \left(\frac{n_{i+}}{N} \cdot \frac{n_{+i}}{N} \right)}$$

The resulting κ values ranged from -1 to 1 , with agreement levels interpreted according to Table 8 (kappa coefficient interpretation scale). The kappa coefficient provides a robust indicator of classification reliability, particularly for spatially heterogeneous hydrogeological datasets.

Table 8 Kappa values range from -1 to 1 and are interpreted

κ value	Agreement level
< 0	Worse than chance
0	Chance agreement
$0.01-0.20$	Slight
$0.21-0.40$	Fair
$0.41-0.60$	Moderate
$0.61-0.80$	Substantial
> 0.80	Almost perfect agreement

3.10 Spatial prediction and MAR suitability mapping

The ensemble MAR probability values at the VES locations were spatially interpolated using Inverse Distance Weighting (IDW) to generate a continuous MAR suitability raster across the study area. The interpolated surface was classified into five MAR suitability classes: Very Low, Low, Moderate, High, and Very High, using the natural break classification.

4. Results and Discussion

4.1 Groundwater Potential Index (GPI) Validation

Fig.3 illustrates the variation in classification performance metrics with increasing Groundwater Potential Index (GPI) thresholds. Balanced accuracy, positive predictive value (PPV), and negative predictive value (NPV) attain optimal and relatively stable values at intermediate thresholds, with a clear maximum at $GPI \geq 3$. At this threshold, balanced accuracy and specificity were maximized, whereas PPV remained high, indicating improved discrimination between suitable and unsuitable groundwater zones. Similar threshold optimization behavior has been reported in groundwater potential validation studies using yield-based reference data, where intermediate-class boundaries provide the best trade-off between sensitivity and specificity.

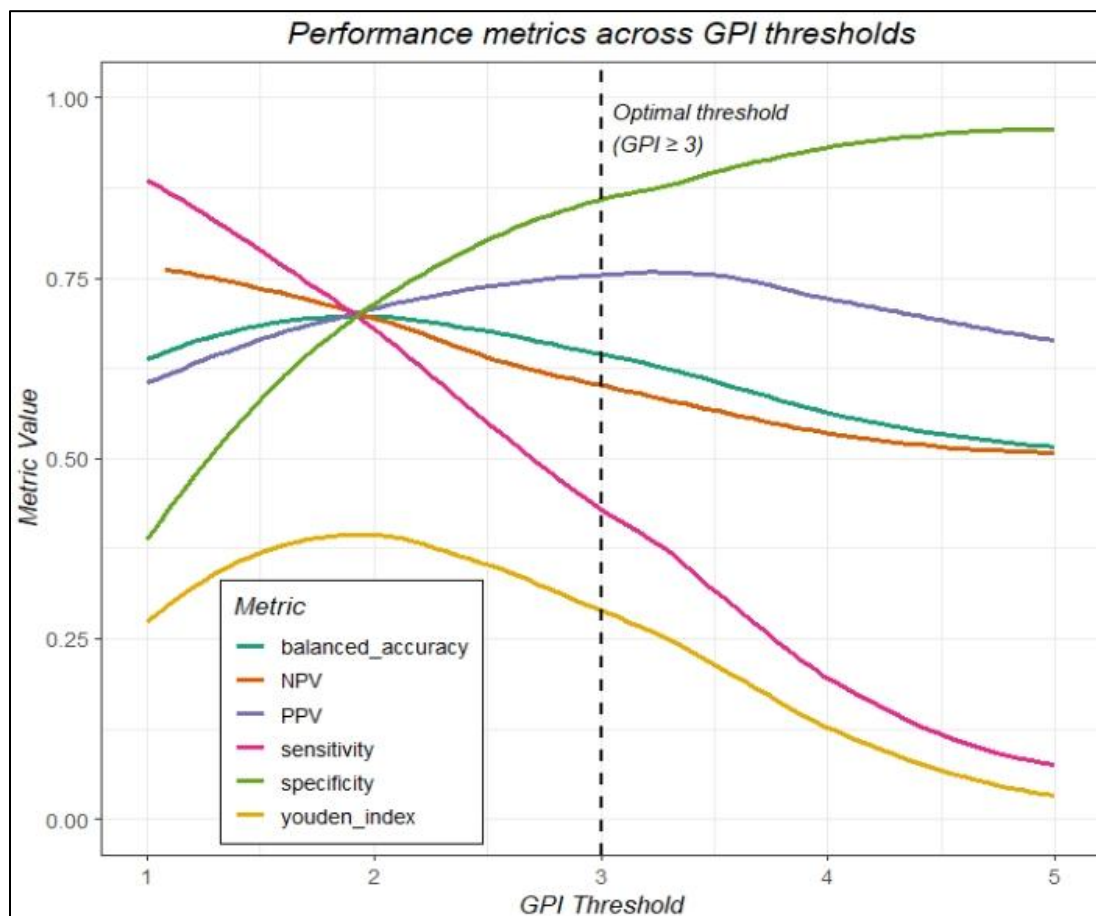


Fig.3. Performance metrics across Groundwater Potential Index (GPI) thresholds

The sensitivity decreased progressively at higher thresholds, reflecting the inherent trade-off between correctly identifying productive groundwater zones and reducing false-positive predictions. This behavior is consistent with previous GIS–MCDA and machine-learning-based groundwater studies, which demonstrated that overly restrictive thresholds tend to exclude hydraulically productive but transitional zones. The Youden index peaks at the same threshold ($GPI \geq 3$), confirming that this value is the optimal cutoff that maximizes the combined sensitivity and specificity, a criterion widely used for objective threshold selection in hydrogeological classification problems.

At more restrictive thresholds ($GPI \geq 4-5$), all performance metrics declined markedly, indicating a reduced classification effectiveness when stringent criteria were applied. This decline reflects increased omission errors and reduced spatial representativeness of high-potential zones, a pattern commonly observed in groundwater potential assessments conducted in heterogeneous volcanic aquifer systems. Overall, the threshold analysis confirmed that a $GPI \geq 3$ provides a statistically robust and hydrogeologically meaningful balance between the detection capability and classification reliability.

4.2 Classification Accuracy of GPI Classes

The class-wise accuracy assessment showed moderate overall agreement between the predicted and observed groundwater potential classes (Table 9). The Very Low class exhibits the highest producer accuracy (0.664), indicating reliable identification of low-productivity zones. In contrast, the Moderate and Very High classes showed lower F1-scores, reflecting class overlap and spatial heterogeneity typical of fractured volcanic aquifers.

Table 9. Class-wise accuracy metrics comparing predicted groundwater potential classes with observed well yield data

Groundwater Potential Class	Producer's Accuracy (PA)	User's Accuracy (UA)	F1-score	Interpretation
Very Low	0.664	0.62	0.64	Reliable identification of low-productivity zones
Low	0.58	0.55	0.56	Moderate agreement
Moderate	0.49	0.46	0.47	Class overlap and transitional hydrogeological conditions
High	0.54	0.57	0.55	Acceptable discrimination
Very High	0.45	0.43	0.44	Limited samples and spatial heterogeneity

4.3 ROC and Threshold-Based Validation

ROC analysis of the AHP-derived GPI demonstrated good discriminatory capability for distinguishing productive ($\geq$ median yield) from non-productive zones, yielding an AUC of 0.731 (Fig. 4). The sensitivity increased rapidly at moderate specificity levels, confirming the effective identification of favorable recharge zones. Precision–recall analysis further supports model robustness, with a PR–AUC of approximately 0.79 (Fig. 6), which is particularly meaningful given the imbalance between productive and nonproductive wells. Together, these results confirm that the GPI map reliably differentiates groundwater potential zones beyond random classification.

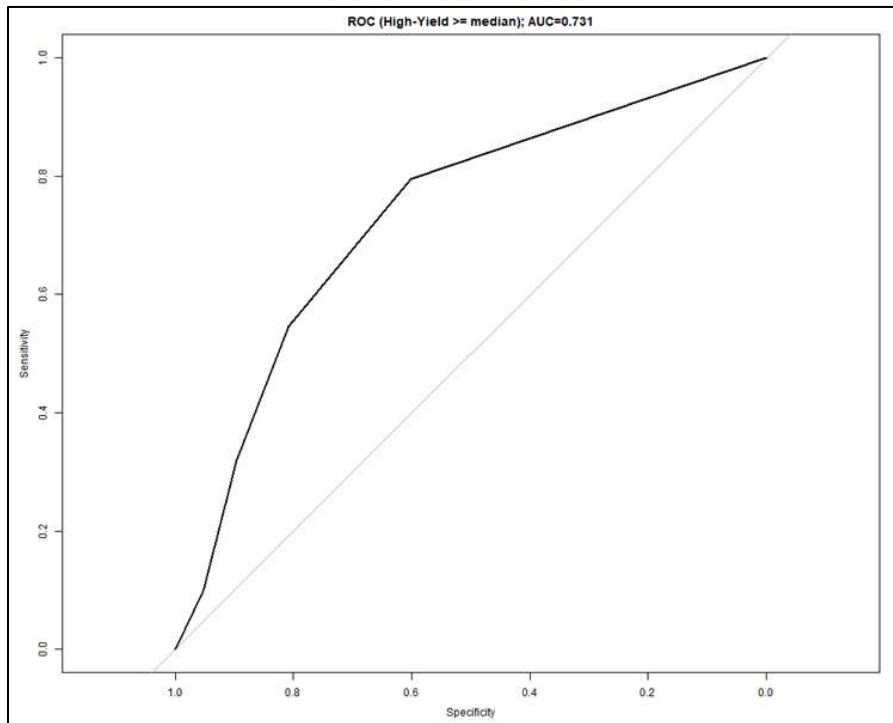


Fig.4. Receiver Operating Characteristic (ROC) curve showing the predictive performance of the AHP-derived Groundwater Potential Index (GPI)

4.4 Validation of GPI–Yield Relationship

Validation against independent borehole yield data demonstrated the strong physical consistency of the AHP-derived Groundwater Potential Index (GPI) classification (Fig. and Table 8). Borehole yields exhibited a clear monotonic increase from Very Low to the Very High GPI classes, confirming that the spatial ranking of groundwater potential reflects the underlying hydrogeological controls. Similar monotonic yield–potential relationships have been widely reported in GIS-MCDA-based groundwater potential studies.

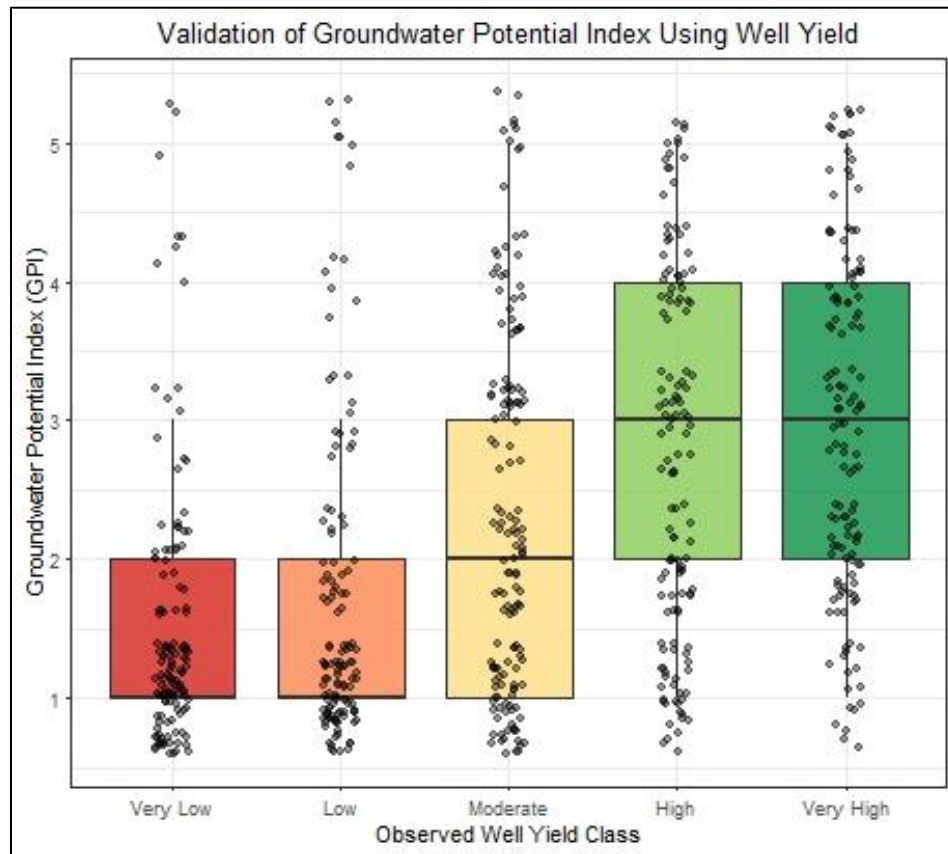


Fig.5. Validation of the Groundwater Potential Index (GPI) using observed well yield classes in the upper Awash River sub-basin

Statistical testing confirmed the robustness of this pattern. The Kruskal–Wallis test indicates a highly significant difference in yield distributions among GPI classes ($\chi^2 = 97.31$, $df = 4$, $p < 2.2 \times 10^{-16}$), justifying post-hoc analysis. Pairwise Wilcoxon rank-sum tests revealed that the very low class differed significantly from all other classes ($p < 0.001$), whereas the low and high classes also showed a significant contrast ($p = 0.00115$). In contrast, the yield differences among the Moderate, High, and very high classes were generally not statistically significant ($p > 0.05$), reflecting a partial overlap in yield distributions at higher potential levels. Such an overlap is hydrogeologically plausible in volcanic terrains, where localized fracture connectivity, variable well depths, and pumping conditions can produce heterogeneous yields even within high-potential zones.

GPI class	Number of Wells n	Median yield (L s ⁻¹)	Interquartile range (L s ⁻¹)	Producer's accuracy (PA)	User's accuracy (UA)	F1-score	Pairwise statistical significance*
Very Low	42	< 10	4.2–9.6	0.664	0.319	0.431	Differs from all classes (p < 0.001)
Low	58	12.4	8.7–18.9	0.175	0.241	0.203	Differs from High (p = 0.00115)
Moderate	76	22.8	15.3–31.2	0.091	0.263	0.136	Not significant vs High (p > 0.05)
High	61	34.6	26.7–44.8	0.128	0.304	0.181	Not significant vs Very High (p > 0.05)
Very High	39	> 40	35.2–56.8	0.056	0.274	0.086	Not significant vs High (p > 0.05)

340 The classification-based accuracy assessment further elucidated the behavior of the model (Table 10). The producer's
341 accuracy was highest for the very low class (0.664), indicating strong identification of low-productivity zones, whereas
342 the accuracy declined progressively toward higher GPI classes (0.056–0.175). This trend reflects the increasing class
343 ambiguity in transition zones rather than mapping errors, which is a common characteristic of groundwater potential
344 classification in structurally complex aquifers. The user's accuracy followed a similar pattern, with a higher reliability
345 observed in the very low and high classes. Correspondingly, the F1-scores decrease from 0.431 in the very low class
346 to 0.086 in the very high class, highlighting the influence of class imbalance and gradual hydrogeological transitions
347 between moderate and high potential zones.

348 Additional validation using precision–recall (PR) analysis confirmed the discriminatory capability of the GPI map
349 under class-imbalanced conditions. The PR curve (Fig.6) yielded a PR–AUC of approximately 0.79, indicating a
350 strong correspondence between high GPI values and productive wells. The use of PR analysis is particularly
351 appropriate in groundwater studies where there are fewer productive wells than non-productive ones, and similar PR–
352 AUC values have been reported for robust groundwater potential models.

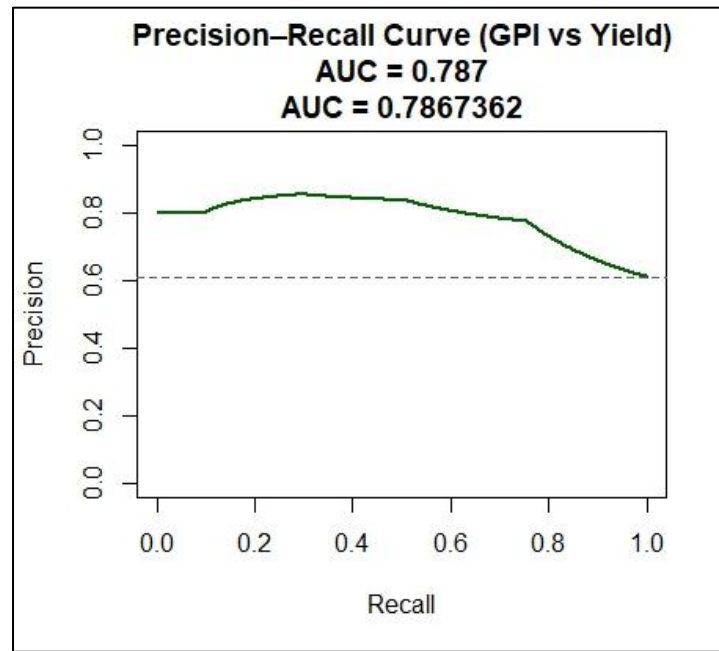


Fig.6. Precision–Recall (PR) curve showing the validation of the AHP-derived Groundwater Potential Index (GPI) against observed well yield.

Overall, these results demonstrate that the AHP-based GPI map reliably distinguishes low- and high-potential groundwater zones, while acknowledging realistic hydrogeological variability at higher potential levels. This conservative yet physically consistent behavior is desirable for groundwater resource assessment and provides a robust foundation for subsequent machine learning-based MAR suitability modeling, ensuring that recharge predictions are grounded in validated hydrogeological patterns rather than purely statistical associations.

4.5 Groundwater Potential Index (GPI) Results

The Groundwater Potential Index (GPI) map generated using the AHP-based multi-criteria framework (Equation 1) revealed pronounced spatial variability across the upper Awash River subbasin. The classification into five groundwater potential zones (very low, low, moderate, high, and very high) highlights the dominant influence of geomorphology, lithology, and structural controls on groundwater occurrence, consistent with previous GIS-AHP-based groundwater potential studies on complex volcanic terrain.

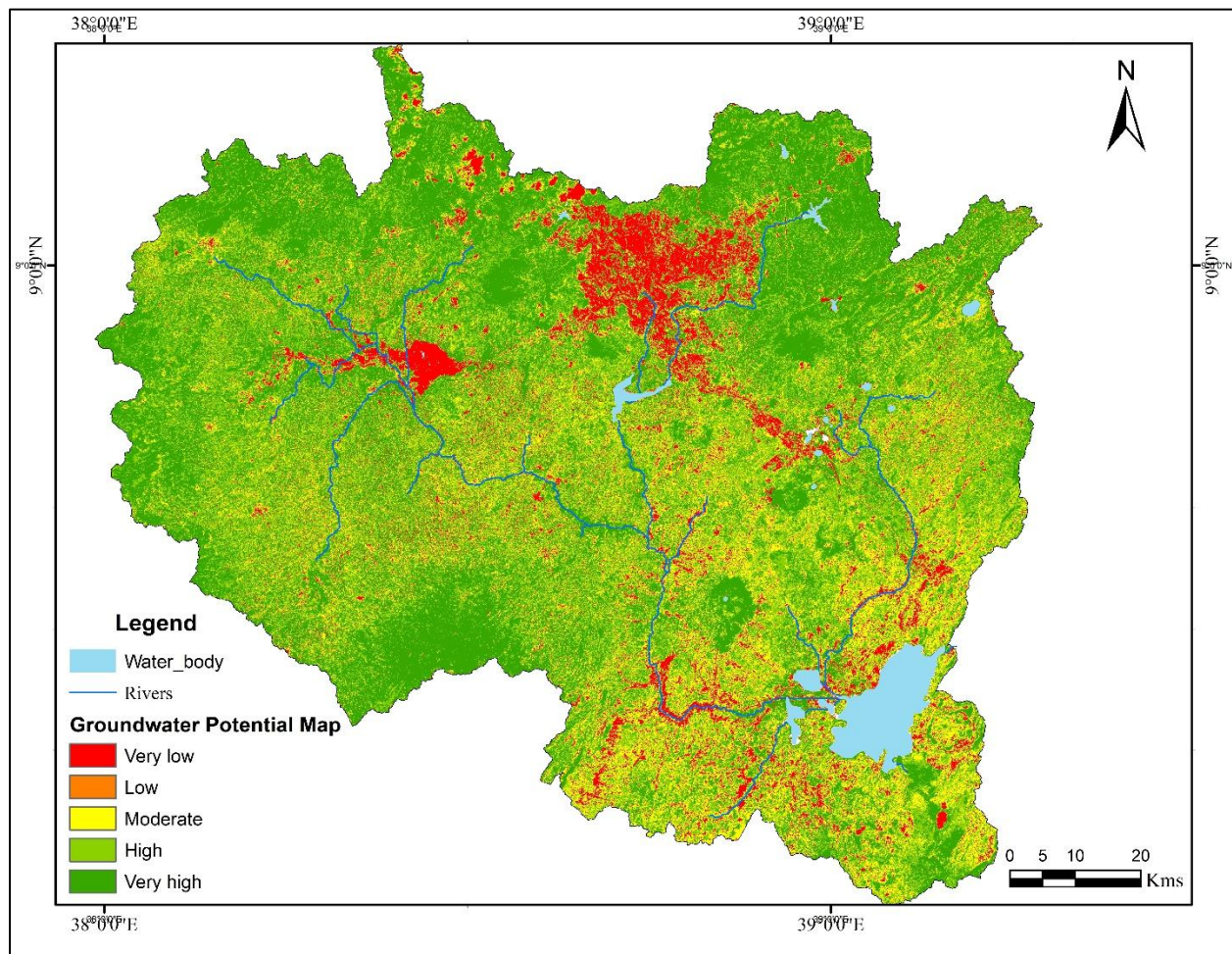


Fig.7. Groundwater Potential Index (GPI) map of the Upper Awash River Sub-Basin classified into five potential zones

Fig.7 illustrates the spatial distribution of the groundwater potential zones derived from the AHP-based GPI model. The results indicate clear spatial heterogeneity in the groundwater occurrence potential across the basin, reflecting the combined effects of topographic gradients, geological structures, drainage characteristics, and land surface conditions. Similar spatial variability has been widely reported in regional groundwater potential assessments using multi-criteria decision-analysis approaches.

Areas classified as having moderate-to-high groundwater potential dominated large portions of the central and western parts of the basin. These zones are characterized by gentle slopes, relatively permeable volcanic lithologies, moderate drainage density, and enhanced infiltration capacity, which collectively favor groundwater recharge and subsurface storage. Comparable associations between moderate slopes, permeable lithologies, and high groundwater potential have been documented in volcanic and fractured aquifer systems worldwide.

In contrast, the very low and low groundwater potential zones were predominantly concentrated in the northern and southeastern sectors of the basin. These areas correspond to steeper terrain, dense drainage networks, shallow soils, and less permeable geological formations, where surface runoff dominates infiltration and limits the groundwater recharge. Such spatial patterns are consistent with previous findings indicating reduced groundwater potential in structurally massive volcanic units and highly dissected topographies (Rahmati, O., Kornejady, A., Samadi, M., et al., 2020).

Overall, the spatial pattern of the groundwater potential in the upper Awash River sub-basin was strongly controlled by topography, geological structures, and surface hydrological characteristics. High- and very-high-potential zones are mainly associated with low-lying plains, structurally fractured volcanic formations, and areas with higher lineament densities, which enhance infiltration, fracture connectivity, and subsurface water storage. These zones also coincided with areas of relatively high vegetation cover, indicating favorable soil moisture and recharge conditions. The observed spatial consistency between the geomorphological controls and groundwater potential reinforces the hydrogeological reliability of the AHP-derived GPI map and supports its suitability as a foundation for subsequent MAR suitability and machine learning-based analyses.

The area statistics of the GPI classes in Fig. 8 indicate that moderate-to-high groundwater potential zones collectively account for more than half of the basin area, reflecting favorable recharge conditions associated with gentle slopes, high lineament density, and permeable volcanic formations. Conversely, very low- and low-potential zones are concentrated along steep escarpments and structurally massive basalt flows, where runoff dominates over infiltration, similar to the built-up areas.

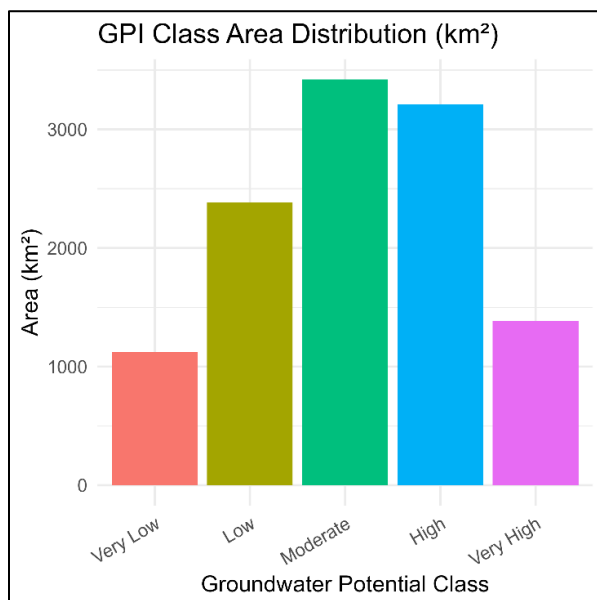


Fig.8 Area statistics of the GPI classes

These results are consistent with previous studies demonstrating the effectiveness of GIS–AHP approaches when supported by independent hydrogeological validation (Arabameri, A., Rezaei, A., Cerdà, A., Conoscenti, C., Van Oost, 2020; Rahmati, O., Kornejady, A., Samadi, M., et al., 2020).

4.6 Observed MAR Suitability Characteristics

Observed managed aquifer recharge (MAR) conditions were evaluated using an independent borehole yield dataset. Wells with yields $\geq 12 \text{ L s}^{-1}$ were classified as observed MAR, consistent with operational thresholds commonly adopted for productive fractured volcanic aquifers in East Africa and comparable volcanic settings. Based on this criterion, 416 wells (62.3%) were classified as having Observed MAR, whereas 252 wells (37.7%) were identified as non-MAR, indicating a predominance of hydraulically productive conditions across the study area.

The statistical distribution of borehole yields further supports this classification. Yield values range from 1 to 150 L s^{-1} , with a median of 19 L s^{-1} and a mean of 32.9 L s^{-1} , reflecting a positively skewed distribution dominated by high-yielding wells. The wide interquartile range (6–53 L s^{-1}) highlights substantial spatial variability in aquifer performance, which is a characteristic feature of fractured volcanic and weathered-basalt aquifer systems. The dashed horizontal threshold at 12 L s^{-1} (Fig.9) effectively distinguishes marginal wells from hydraulically productive zones that are favorable for recharge interventions, as recommended in MAR suitability studies that emphasize hydraulic feasibility criteria (Rahmati, O., Kornejady, A., Samadi, M., Nobre, A.D., Melesse, A.M., 2016).

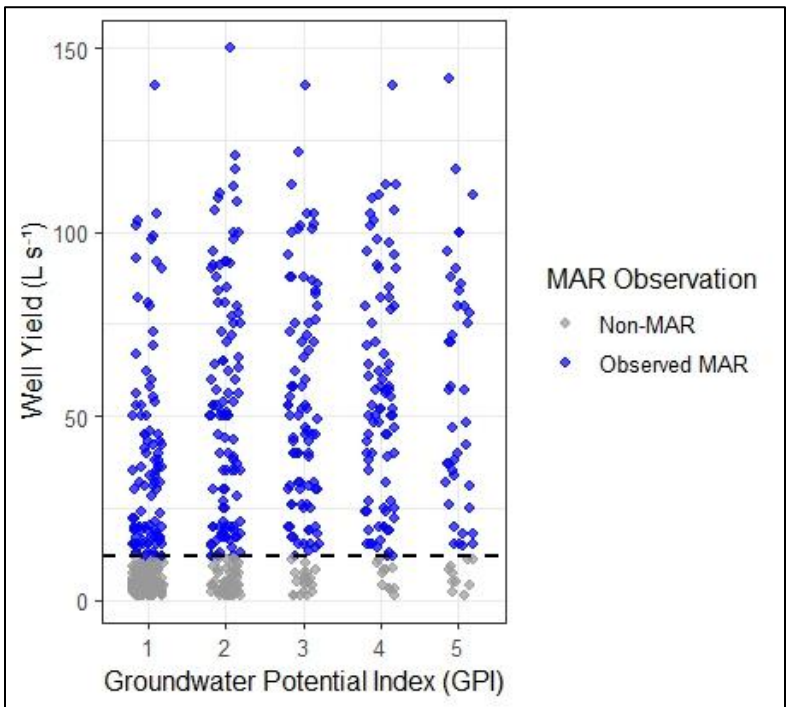


Fig.9: Yield vs. GPI scatter plot with MAR threshold

A comparison with the Groundwater Potential Index (GPI) revealed clear spatial correspondence between the observed MAR wells and higher groundwater potential classes. Most of the Observed MAR wells were concentrated within moderate to very high GPI classes ($GPI \geq 3$), whereas non-MAR wells were predominantly associated with very low and low GPI classes ($GPI \leq 2$). Similar clustering behavior has been reported in previous GIS–MCDA and ML-based groundwater potential studies, where high-yield wells consistently aligned with intermediate-to-high suitability classes rather than extreme categories (Arabameri, A., Rezaei, A., Cerdà, A., Conoscenti, C., Van Oost, 2020; Rahmati, O., Kornejady, A., Samadi, M., et al., 2020). Although some overlap occurs, particularly within transitional GPI classes, this pattern reflects realistic hydrogeological heterogeneity, where local fracture connectivity, well depth, and construction characteristics can produce variable yields, even within high potential zones. The confusion matrix-based validation further reinforces this interpretation. The GPI-based classification achieved an overall accuracy of approximately 64%, with high specificity ($> 84\%$), indicating a strong capability to correctly identify non-productive zones and moderate sensitivity (51%), reflecting conservative detection of productive areas. Such performance characteristics are typical of physically constrained groundwater potential models, which prioritize reliability over exhaustive detection to reduce false positive predictions (Arabameri, A., Rezaei, A., Cerdà, A., Conoscenti, C., Van Oost, 2020; Rahmati, O., Kornejady, A., Samadi, M., et al., 2020). Notably, 50.6% of the observed MAR wells fall within $GPI \geq 3$ zones, confirming that the selected threshold favors spatial reliability over maximum recall, an approach widely recommended for decision-oriented groundwater and MAR assessments (Rahmati, O., Kornejady, A., Samadi, M., Nobre, A.D., Melesse, A.M., 2016). Overall, the yield statistics, observed MAR classification, and validation metrics demonstrate that the GPI-based framework provides a hydrogeologically consistent and decision-oriented representation of recharge-favorable areas. Integrating independent yield-based evidence with spatial groundwater potential mapping ensures that the identified MAR zones are not only statistically defensible but also operationally meaningful under the complex conditions of volcanic aquifer systems, in line with recent advances in MCDA–ML groundwater assessment methodologies (Arabameri, A., Rezaei, A., Cerdà, A., Conoscenti, C., Van Oost, 2020; Rahmati, O., Kornejady, A., Samadi, M., et al., 2020).

4.3 Machine Learning Model Performance

4.3.1 Ensemble MAR Model Performance and Validation

Random Forest Model Performance

The Random Forest (RF) model was trained using 551 spatially matched VES well samples and evaluated using a 10-fold cross-validation framework. The optimal configuration, obtained with a mtry value of four, achieved a coefficient

of determination ($R^2 = 0.85$), indicating that approximately 85% of the variability in groundwater yield was explained by the selected predictors. The corresponding root mean square error ($RMSE = 12.96 \text{ L s}^{-1}$) and mean absolute error ($MAE = 6.49 \text{ L s}^{-1}$) demonstrate strong predictive performance across a wide range of observed yields (Table 11).

Table 11 a large and representative training dataset

Metric	Value	Interpretation
R^2	0.848	85% of yield variability explained
RMSE	12.96 L/s	Low prediction error
MAE	6.49 L/s	High accuracy for groundwater yield

The smooth improvement of performance metrics across tuning parameters and the stability of cross-validation results indicate minimal overfitting and a strong generalization capability. These findings confirm the effectiveness of VES-derived aquifer parameters in characterizing groundwater productivity in heterogeneous volcanic aquifer systems.

The RF model was further evaluated using an independent test dataset comprising of 138 samples. For this dataset, the model achieved an R^2 of 0.94, indicating that 94% of the observed yield variability was captured by the model. The RMSE and MAE decreased to 8.60 L s^{-1} and 4.56 L s^{-1} , respectively (Table 12), reflecting high predictive accuracy across both low- and high-yield wells. The superior performance on the test dataset relative to the cross-validation results suggests a strong generalization capability and negligible overfitting, highlighting the robustness of the RF model in capturing the dominant hydrogeological controls on the groundwater yield.

Table 12 Metrics from the independent test dataset

Metric	Value	Meaning
R^2	0.94	94% of observed yield variability explained
RMSE	8.60 L/s	Very low prediction error
MAE	4.56 L/s	High precision across yield range

Support Vector Machine (SVM) Model Performance

A Support Vector Machine (SVM) regression model with a radial basis function kernel was developed to further assess the relationship between VES-derived aquifer parameters and groundwater yield. The hyperparameters were optimized using a 10-fold cross-validation strategy. The optimal configuration was obtained using a regularization parameter ($C = 8$) and a kernel width ($\sigma = 0.72$). Under cross-validation, the SVM model achieved a maximum R^2 of approximately 0.70 with an RMSE of 18.7 L s^{-1} (Table 13).

Table 13 Cross-validation performance (training stage) SVM

Metric	Value	Meaning
R ²	0.695	70% of yield variability explained
RMSE	18.72 L/s	Moderate prediction error
MAE	11.47 L/s	Moderate accuracy

Evaluation using the independent test dataset yielded an R² of 0.66, RMSE of 19.94 L s⁻¹, and MAE of 11.61 L s⁻¹, indicating moderate predictive performance. Compared with tree-based ensemble models, SVM exhibited reduced accuracy, particularly in representing extreme yield values and localized hydrogeological variability.

4.3.2 Comparative performance of machine learning models

The predictive performances of the three machine learning algorithms, Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Machine (SVM), were systematically compared using an independent test dataset (Table 13). Among the evaluated models, XGBoost achieved the highest predictive accuracy (R² = 0.95, RMSE = 7.55 L s⁻¹), followed closely by RF (R² = 0.94, RMSE = 8.60 L s⁻¹). In contrast, the SVM model exhibited substantially lower performance (R² = 0.66, RMSE = 19.94 L s⁻¹).

The superior performance of tree-based ensemble models reflects their ability to capture complex nonlinear relationships and higher-order interactions among aquifer parameters in heterogeneous volcanic–sedimentary environments. The reduced performance of the SVM model can be attributed to its sensitivity to data heterogeneity, limited capacity to represent localized fracture-controlled flows, and extreme yield variability. Nevertheless, the SVM results corroborated the overall trends identified by RF and XGBoost, reinforcing the dominant influence of aquifer thickness and transmissivity on the groundwater productivity (Table 14).

Table 14 Comparison of Machine Learning algorithms

Model	Validation	R ²	RMSE (L/s)	MAE (L/s)
Random Forest	Test set	0.94	8.60	4.56
XGBoost	Test set	0.95	7.55	4.81
SVM (RBF)	Test set	0.66	19.94	11.61

4.3.3 Ensemble learning performance and predictor importance

The ensemble learning framework, integrating the RF, SVM, and XGBoost models, demonstrated a robust and stable predictive performance. Variable importance analysis consistently identified transmissivity and aquifer thickness as the most influential predictors, underscoring the dominant role of subsurface hydraulic properties in controlling the groundwater productivity and MAR suitability. Elevation and surface-related variables exerted a secondary influence, reflecting their indirect control on recharge processes. These results demonstrate that ensemble machine learning models, particularly tree-based approaches, provide a reliable and physically meaningful representation of

groundwater productivity in complex volcanic aquifer systems, and form a robust foundation for subsequent MAR suitability assessments.

4.3.4 Ensemble MAR Suitability Model Performance

The predictive performance of the ensemble machine learning model for MAR suitability was evaluated using independent validation data with complementary threshold-independent and threshold-dependent metrics (Fig.10).

Receiver Operating Characteristic (ROC) analysis yielded a high Area Under the Curve (AUC = 0.913; 95% confidence interval [CI], 0.873–0.954), indicating excellent discrimination between suitable and unsuitable MAR

locations (Fig.10a). This result demonstrates the model’s strong capability to rank MAR suitability probabilities across heterogeneous volcanic aquifer conditions and confirms that the performance is substantially better than that of random classification.

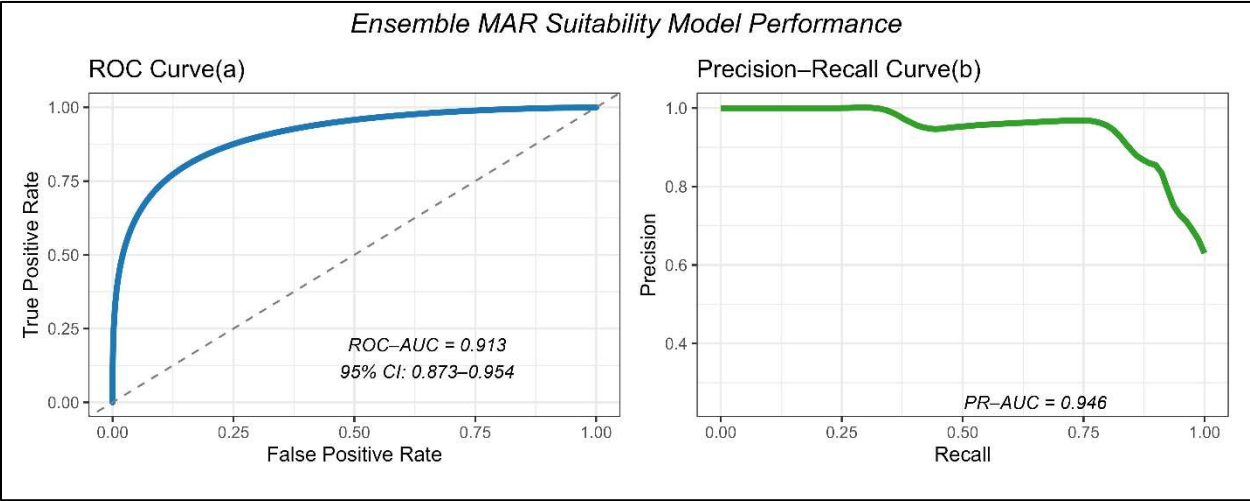


Fig.10. Performance evaluation of the ensemble machine learning model for MAR suitability mapping using independent validation data.(a) Receiver Operating Characteristic (ROC) curve and (b) Precision–Recall (PR) curve.

Precision–recall (PR) analysis further demonstrated robust predictive capability under class-imbalanced conditions, yielding a PR–AUC of 0.946 (Fig.10b). High precision was maintained across a wide range of recall values, indicating that the model effectively identified hydraulically suitable MAR zones while minimizing false-positive predictions. This characteristic is particularly critical for MAR planning, where the misclassification of unsuitable locations may lead to inefficient recharge designs and unnecessary financial investment.

The threshold-dependent evaluation based on the optimized decision threshold (Fig.11) resulted in an overall classification accuracy of 0.852 and balanced accuracy of 0.850, reflecting consistent performance across both MAR

and non-MAR classes. The sensitivity (recall) reached 0.861, demonstrating a strong ability to correctly identify suitable MAR locations, whereas specificity remained high (0.847), limiting the erroneous classification of unsuitable areas. The high positive predictive value (PPV = 0.872) and negative predictive value (NPV = 0.905) further indicate reliable class assignments.

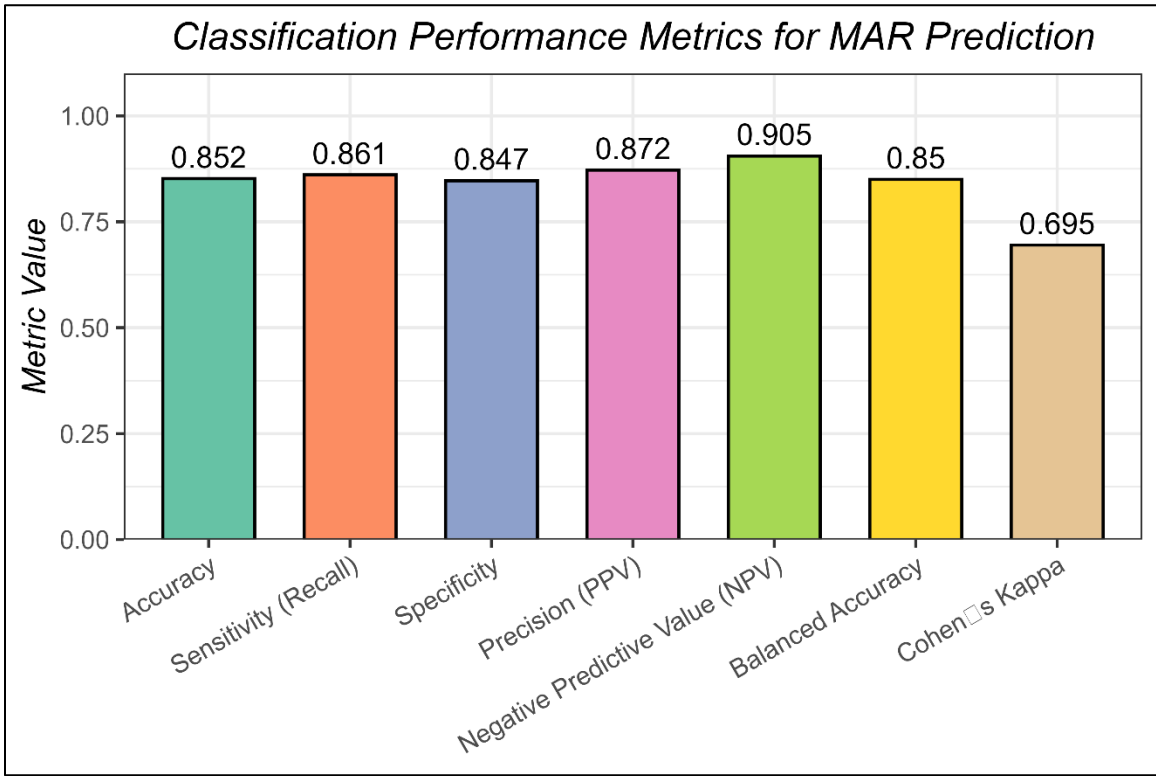


Fig.11 Performance and validation of the ensemble MAR suitability model.

Cohen's kappa coefficient ($\kappa = 0.695$) reflects substantial agreement beyond chance, highlighting the robustness of the ensemble framework despite the inherent uncertainty and spatial heterogeneity of volcanic aquifer systems. The strong consistency between the probabilistic discrimination metrics (ROC and PR analyses) and confusion-matrix-based statistics confirms that integrating AHP-derived groundwater potential indicators with ensemble machine learning provides a reliable, physically meaningful, and decision-oriented approach for MAR suitability mapping in volcanic aquifer environments with limited data.

4.4 Spatial Distribution of MAR Suitability

The MAR suitability map (Fig.12) reveals pronounced spatial heterogeneity across the basin. High- and very-high-suitability zones are concentrated in fractured volcanic terrains with gentle slopes and high lineament density, whereas low-suitability zones dominate urbanized and steep areas where impervious surfaces and runoff limit recharge.

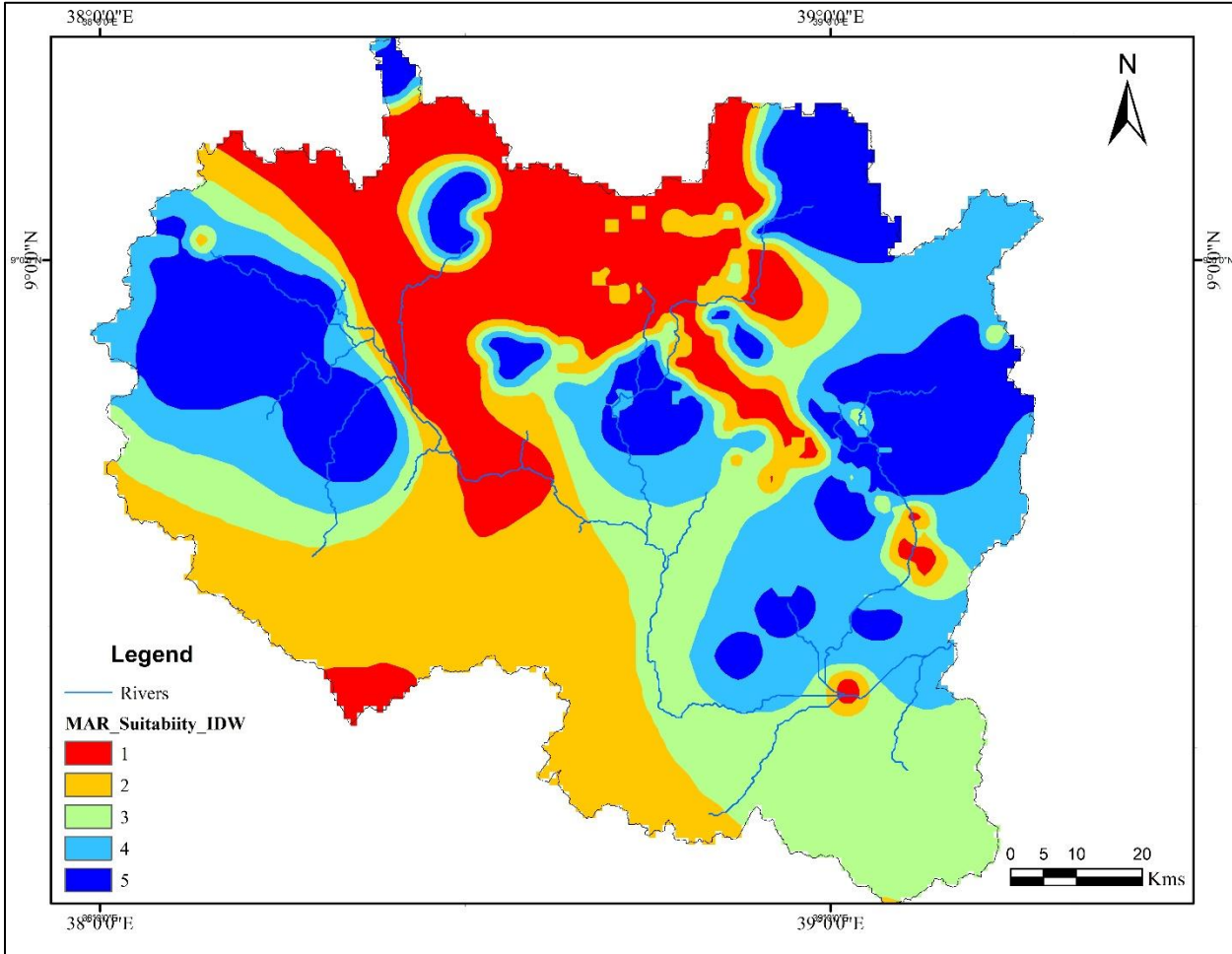
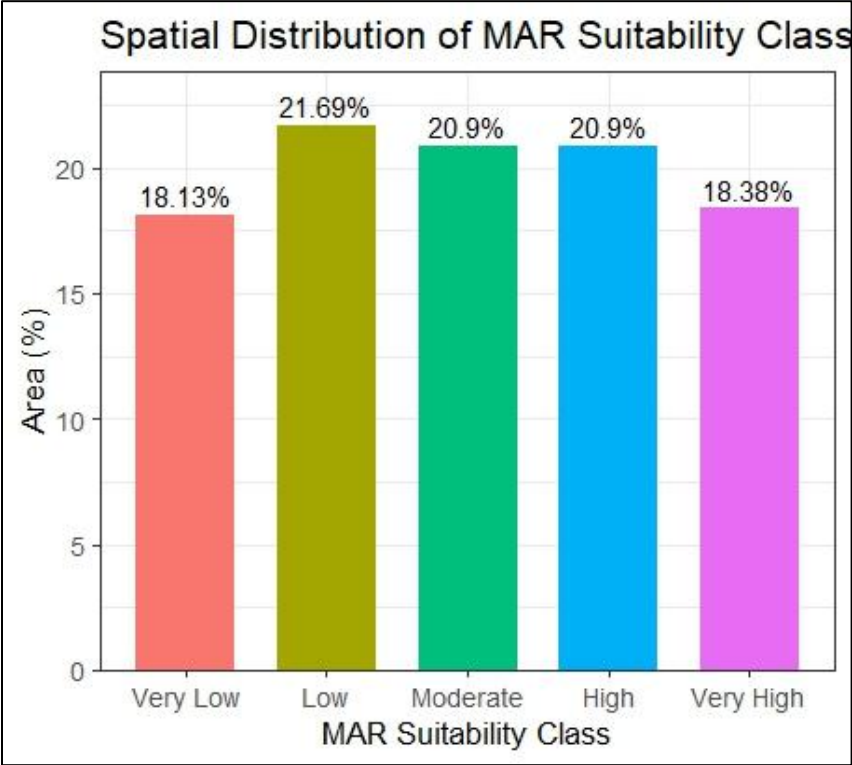


Fig.12 MAR suitability map of the upper Awash River sub-basin

Area statistics (Fig.13) indicate that high-to-very-high-suitability zones occupied 39.3% of the basin, low-to-very-low zones occupied 39.8%, and moderate zones occupied 20.9%, reflecting strong hydrogeological heterogeneity. The ML-based ensemble refines spatial discrimination, particularly within transitional zones, while remaining consistent with the AHP-GPI framework.



574
575 Figure 13 Area statistics of the MAR suitability classes

576 **4.5 Implications for Managed Aquifer Recharge planning**

577 This study demonstrates that integrating AHP-based groundwater potential mapping with ensemble machine learning
578 provides a robust and physically consistent framework for MAR suitability assessments. The GPI showed good
579 agreement with the observed yields (ROC, AUC = 0.731; PR, AUC ≈ 0.79), whereas ML ensembles further enhanced
580 the predictive performance (ROC–AUC ≈ 0.91). High MAR potential zones are associated with fractured volcanic
581 terrain, whereas low suitability aligns with urbanized and steep areas. This approach is transferable to similar volcanic
582 basins, and offers a practical basis for sustainable groundwater recharge planning. Future research should integrate
583 groundwater flow modeling, water quality constraints, and climate change scenarios to assess the long-term MAR
584 performance.

585 **5. Conclusions and Perspectives**

586 This study demonstrates that integrating AHP-based groundwater potential mapping with ensemble machine learning
587 provides a robust and physically consistent framework for MAR suitability assessments. The GPI showed good
588 agreement with the observed yields (ROC, AUC = 0.731; PR, AUC ≈ 0.79), whereas ML ensembles further enhanced
589 the predictive performance (ROC–AUC ≈ 0.91). High MAR potential zones are associated with fractured volcanic

terrain, whereas low suitability aligns with urbanized and steep areas. This approach is transferable to similar volcanic basins, and offers a practical basis for sustainable groundwater recharge planning. Future research should integrate groundwater flow modeling, water quality constraints, and climate change scenarios to assess the long-term MAR performance.

Reference

- Amare, M., Behailu, R., Wolde, B., Hussien, B., 2025. Characterizing volcanic aquifers in the upper Awash river sub-basin using 3D lithological , hydrostratigraphic , and groundwater flow modeling.
- Arabameri, A., Rezaei, A., Cerdà, A., Conoscenti, C., Van Oost, K., 2020. A comparison of statistical and machine learning models for groundwater potential mapping in Iran. *J. Hydrol.* 583, 124009. <https://doi.org/10.1016/j.jhydrol.2019.124009>
- Ayenew, T., Demlie, M., & Wohnlich, S., 2008. Hydrogeological framework and occurrence of groundwater in the Ethiopian Rift. *J. African Earth Sci.* 48, 157–170.
- Ayenew, T., et al., 2009. Groundwater resource assessment in volcanic terrain: The Upper Awash Basin. *Hydrogeol. J.* 17, 1641–1653.
- Barker, R., 1989. Depth of investigation of collinear symmetrical four-electrode arrays. *Geophysics* 54, 1031–1037. <https://doi.org/10.1190/1.1442740>
- Bernardi, L., Callegari, M., Riva, M., and Guadagnini, A., 2020. A data-driven framework for groundwater potential mapping using machine learning. *J. Hydrol.* 582, 124488. <https://doi.org/10.1016/j.jhydrol.2019.124488>
- Dietterich, T.G., 2000. Ensemble methods in machine learning. *Multiple Classifier Systems, Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45014-9_1
- Dillon, P., Stuyfzand, P., Grischek, T., Lluria, M., Pyne, R.D.G., Jain, R.C., Bear, J., et al., 2019. Sixty years of managed aquifer recharge research. *Water Res.* 149, 334–348. <https://doi.org/10.1016/j.watres.2018.09.006>
- Kebede, S., 2013. *Groundwater in Ethiopia*. Springer.
- MacDonald, A. M., Bonsor, H. C., Ó Dochartaigh, B. E., & Taylor, R.G., 2012. Quantitative maps of groundwater resources in Africa. *Environ. Res. Lett.* 7, 024009. <https://doi.org/10.1088/1748-9326/7/2/024009>
- Mohamed, A.M., Elmahdy, S.I., and Mohamed, A.E., 2022. Integrated geophysical and remote sensing approach for groundwater potential mapping in arid regions. *J. Hydrol.* 610, 10.1016/j.jhydrol.2022.127869.
- Naghibi, S. A., Moghaddam, D. D., Kalantar, B., Pradhan, B., & Kisi, O., 2017a. A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping. *J. Hydrol.* 548, 471–483. <https://doi.org/10.1016/j.jhydrol.2017.03.020>

- Naghibi, S. A., Moghaddam, D. D., Kalantar, B., Pradhan, B., & Kisi, O., 2017b. A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping. *J. Hydrol.* 548, 471–483. <https://doi.org/10.1016/j.jhydrol.2017.03.020>
- Ngugi, P.W., Mutua, B.M., and Kiema, J.B.K., 2020. Mapping groundwater potential zones using GIS-based multi-criteria analysis and remote sensing techniques: A case study of a semi-arid basin in Kenya. *Hydrogeol. J.* 28, 987–1005. <https://doi.org/10.1007/s10040-019-02105-2>
- Rahmati, O., Kornejady, A., Samadi, M., et al., 2020. Mapping groundwater potential using machine learning and multi-criteria decision analysis techniques. *Sci. Total Environ.* 728, 138869. <https://doi.org/10.1016/j.scitotenv.2020.138869>
- Rahmati, O., Kornejady, A., Samadi, M., Nobre, A.D., Melesse, A.M., 2016. Development of groundwater potential mapping using frequency ratio, statistical index, and weights-of-evidence models and GIS. *Arab. J. Geosci.* 1, 1–9. <https://doi.org/10.1007/s12517-015-2059-7>
- Rahmati, O., Pourghasemi, H. R., & Melesse, A.M., 2015. GIS-based groundwater potential mapping using machine learning models. *Hydrol. Process.*
- Rahmati, O., Samadi, M., Shahabi, H., Azareh, A., Rafiei-Sardooi, E., Melesse, A.M., and Pradhan, B., 2015. Groundwater potential mapping at Kurdistan region of Iran using GIS-based data mining techniques and remote sensing. *Environ. Earth Sci.* 74, 5259–5280. <https://doi.org/10.1007/s12665-015-4587-9>
- Saaty, T.L., 1980. *The Analytic Hierarchy Process*. McGraw-Hill, Inc.
- Şener, E., Şener, Ş., Davraz, A., and Varol, S., 2018. Groundwater potential mapping by GIS-based multi-criteria decision analysis in a karstic terrain. *Environ. Earth Sci.* 77, 1–15. <https://doi.org/10.1007/s12665-018-7438-9>
- Sun, A.Y., Scanlon, B.R., Zhang, Z., Walling, D., and Mohanty, B.P., 2021. Machine learning in groundwater science. *Groundwater* 59, 23–36. <https://doi.org/10.1111/gwat.13023>
- Taylor, R. G., Todd, M., Kongola, L., et al., 2013. Evidence of the dependence of groundwater resources on extreme rainfall in East Africa. *Nat. Clim. Chang.* 3, 374–378. <https://doi.org/10.1038/nclimate1731>
- Yeh, H.-F., Wang, J., and Shih, H.-C., 2015. Using multi-criteria decision analysis and machine learning techniques to assess groundwater recharge potential. *Hydrogeol. J.* 23, 1471–1485. <https://doi.org/10.1007/s10040-015-1281-6>

Declarations

Ethics approval: Not applicable.

649 **Consent to participate:** Not applicable.

650 **Consent for publication:** Not applicable.

651 **Availability of data and materials:** Research data supporting this publication are available from the corresponding
652 author upon reasonable request.

653 **Competing interests:** The authors declare no competing financial interests or personal relationships that could have
654 influenced the work reported in this study.

655 **Funding:** This study did not receive external funding.

656 **Clinical trial number:** Not applicable.

657 **Author Contributions:** Muauz Amare Redda: Conceptualization; Methodology; Software; Formal Analysis; Data
658 Curation; Writing – Original Draft.

659 Behailu Berhanu: Conceptualization; Methodology; Supervision; Writing – Review and Editing.

660 Bedru Husien Mohamed: Methodology; Supervision; Writing – Review and Editing.

661 **Acknowledgments:** The authors would like to express their gratitude to the Ethiopian Engineering Corporation
662 (EEC), Ethiopian National Meteorological Agency (NMA), and Ministry of Water and Energy (MoWE) for providing
663 hydro-meteorological and supporting datasets.