International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

Artificial Intelligence and Multivariate Analysis of Sodium Flushing and Fluoride Signatures in the Merti Transboundary Aquifer: Hydrochemical Evidence of Climate Stress and Adaptation Financing Implications

Dr. Meshack Owira Amimo[1], Ms. Agatha Njuguna[2], Mr. Daniel Ombaka[3]

[1]Hydrogeologist & Research Scholar, IIC University of Technology (Phnom Penh), Cambodia
[2]Hydrogeologist & HoA Coordinator, Water Resources Authority (WRA), Nairobi
[3]Analytical Chemist & Water Quality and Pollution Control Officer, Water Resources Authority Nairobi

Correspondence Author: Dr. Meshack Owira Amimo

(cc) BY

Abstract: The Merti Aquifer, a confined transboundary system spanning Isiolo, Garissa, and Wajir counties in Kenya and extending into Somalia's Jubbada Hoose region, sustains over one million people and vast livestock populations in one of the most arid environments in the Horn of Africa. Despite its critical importance, the hydrochemical evolution of the aquifer under climate stress and over-abstraction remains poorly quantified. This study integrates Principal Component Analysis (PCA) with a machine learning (ML) benchmarking framework to model Sodium dynamics and assess climate change impacts. Using a hydrochemical dataset of 31 parameters, subsequently refined to 14 key predictors, we evaluated the latent structure of groundwater chemistry and compared six base regression algorithms: Random Forest, Decision Tree, Back-Propagation Neural Network (BPNN), Ridge Regression, Lasso Regression, and k-Nearest Neighbors, alongside a Stacking Ensemble. PCA revealed two dominant orthogonal processes: PC1 (36.1% variance) represents evaporative concentration and salinization driven by Sodium, Chloride, and TDS; PC2 (18.3% variance) captures volcanic geogenic signatures including Fluoride and Sulfide. The orthogonal separation of Sodium and Fluoride confirms distinct origins—Sodium is driven by climate-induced stagnation in the lowlands, while Fluoride originates from Mt. Kenya's volcanic highlands. The Stacking Ensemble achieved exceptional predictive accuracy for Sodium ($R^2 = 0.9856$, RMSE = 164.07), closely followed by BPNN ($R^2 = 0.9793$). Cross-validated R-scores identified Ridge ($R = 0.98$) and BPNN ($R = 0.97$) as the most stable predictors. We conclude that Sodium and Chloride serve as reliable conservative proxies for climate-induced salinization, and that ML-based diagnostic dashboards can inform targeted climate adaptation funding, including managed aquifer recharge and transboundary water governance.

Keywords: Merti Aquifer; Sodium dynamics; Principal Component Analysis; Machine Learning; Climate Change; Transboundary Groundwater; Kenya.

I. INTRODUCTION

The Merti Aquifer is the largest and most economically significant groundwater resource in northern Kenya, extending approximately 73,713 km² from Sericho and Modogashe in Isiolo County through Habaswein and Dadaab in Garissa, to Wajir and across the border into Somalia's Jubbada Hoose region [1]. It provides the sole reliable water source for over one million pastoralists, agro-pastoralists, and refugees, including the Dadaab refugee complex, which hosts over 350,000 people. The aquifer is a confined sedimentary system hosted within the Anza Graben, a deep tectonic basin filled with Mesozoic to Quaternary formations [2]. Groundwater is typically encountered at depths of 110–180 meters below ground level, with the productive "Merti beds" comprising friable to semi-consolidated gravels, sands, silts, and clays.

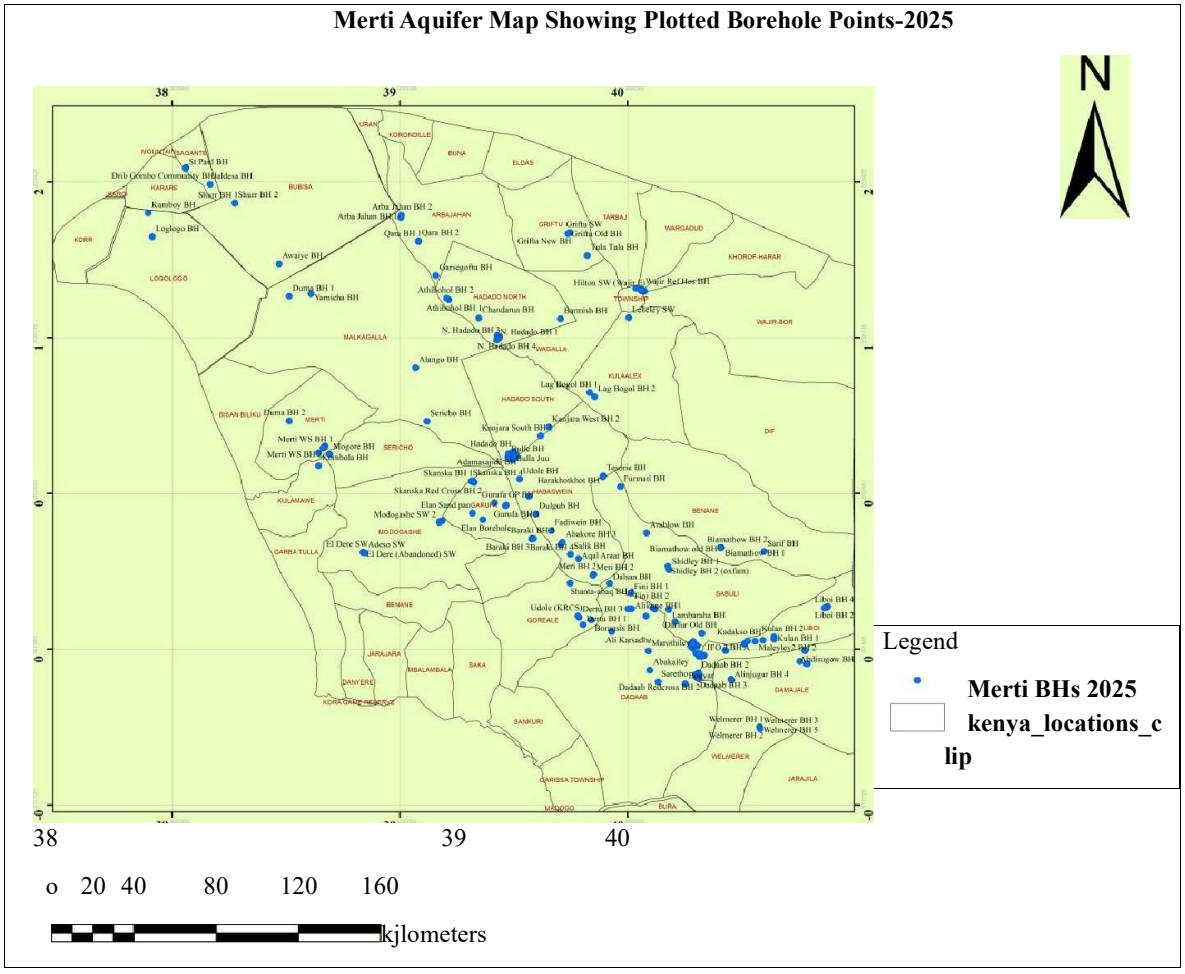
The hydrogeological significance of the Merti Aquifer lies in its recharge mechanism, which is fundamentally decoupled from local precipitation. While rainfall across the aquifer's footprint rarely exceeds 300 mm/year, the system is recharged via focused infiltration along the Ewaso Ng'iro and Lagh Dera drainage systems, which originate in the highlands of Mt. Kenya and the Aberdare Range [5]. This transboundary flow path—from high-elevation volcanic terrain (>2,000 m ASL) descending to the arid plains (90–400 m ASL)—induces a predictable hydrochemical evolution: fresh, dilute recharge waters progressively acquire dissolved solids through water-rock interaction, evaporation, and mixing with deeper saline formations.

Climate change threatens this delicate balance. Prolonged droughts and erratic rainfall in the Kenyan highlands have reduced the frequency and magnitude of flushing events that historically pushed salts out of the central aquifer corridor toward Somalia [3]. Simultaneously, over-abstraction—driven by population growth, refugee influxes, and livestock demand—has lowered water tables and accelerated up-coning of saline water from deeper formations. The result is a dual crisis: declining water availability and deteriorating water quality, manifesting most visibly as rising Sodium and Chloride concentrations.

This study addresses three interrelated questions. First, what is the latent hydrochemical structure of the Merti Aquifer, and how does it distinguish between geogenic (volcanic) and anthropogenic-climatic (evaporative) salinity sources? Second, can machine learning algorithms reliably predict Sodium concentrations from other hydrochemical parameters, thereby enabling low-cost monitoring? Third, how can these findings inform climate adaptation financing and transboundary water governance?

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

II. GEOGRAPHY AND HYDROGEOLOGY OF THE MERTI AQUIFER



Map 1: the wells drilled in the Merti Aquifer as captured in a study between 2018 and 2025

A. Geographic Extent and Administrative Coverage

The Merti Aquifer spans four Kenyan counties—Marsabit, Isiolo, Garissa, and Wajir—and extends into Somalia's Jubbada Hoose region. The aquifer's central axis follows the Ewaso Ng'iro/Lagh Dera drainage corridor, which flows eastward from the highlands near Mount Kenya through Sericho, Merti, Habaswein, and Dadaab before terminating in the Somali lowlands near Afmadow and Goobweyn. The western boundary is marked by the volcanic slopes of Mt. Kenya and Mt. Marsabit (both exceeding 1,500 m ASL), while the eastern extent descends to approximately 400 m ASL near the Somali border.

The climate is classified as hot arid and semi-arid (BWh in the Köppen system). Rainfall is below 300 mm/year across most of the aquifer's footprint, with potential evaporation exceeding 2,000 mm/year. Surface water is ephemeral, flowing only during exceptional flood events associated with El Niño cycles. The Ewaso Ng'iro River, historically a perennial source, now typically dries up at Merti town in most years, with flow reaching Liboi only during major floods [5].

B. Geological Framework and Aquifer Architecture

The Merti Aquifer is hosted within the Anza Graben, a deep sedimentary basin filled with Cretaceous, Paleogene, and Neogene sediments overlying Precambrian basement [2]. The upper 600 meters of the basin comprise intercalated alluvial and peri-marine sediments of Miocene and Pliocene age. The productive unit—the "Merti beds"—is a late Miocene horizon of coarse sands and gravels located at depths of 90–150 meters, with thicknesses of 20–40 meters. This unit extends over 200 km from Habaswein to Dadaab and Liboi.

The aquifer is confined to semi-confined, with storage coefficients ranging from 10^{-5} to 10^{-3} and transmissivity values of 5–50 m²/day in the central corridor. Borehole yields vary significantly: central corridor wells can produce over 40 m³/hr, while peripheral wells yield only 7–16 m³/hr and exhibit poor recovery. The freshwater-saline interface is shallow in low-lying areas (as shallow as 150 m bgl) and deeper in elevated zones (>350 m bgl). This interface is dynamically vulnerable to up-coning under pumping stress [7].

A deeper, potentially fossil aquifer has been postulated at depths exceeding 300 meters within Mesozoic formations, though its hydraulic properties remain unverified. This deep unit may contain significant reserves of paleo-recharge, but its exploitation would carry high risks of irreversible salinization.

C. Recharge Dynamics and the Transboundary Flow Path

Recharge to the Merti Aquifer is dominated by focused infiltration along the Ewaso Ng'iro and Lagh Dera systems, with an estimated annual recharge of approximately 3.3 million cubic meters (MCM) from the Lorian Swamp and Lagh Dera [1]. However, the precise pathways and rates remain poorly constrained. Isotope studies by the IAEA have shown that most groundwater samples cluster near river water composition with minimal evaporative enrichment, suggesting rapid recharge during flood events [5]. Radiocarbon dating from the 1970s indicated that some groundwater was recharged during major floods occurring on timescales of hundreds to thousands of years, implying that large portions of the aquifer contain "fossil" or paleo-recharge.

The transboundary nature of the aquifer is legally significant. A Memorandum of Understanding facilitated by IGAD for the Dhobley–Liboi pilot project was approved by Kenya in 2025 and sent to Somalia for approval, establishing a framework for coordinated development and management [3]. The aquifer's flow terminates in Somalia, meaning that upstream abstraction in Kenya directly affects downstream availability and quality in Somali communities.

III. STRESS ON THE AQUIFER: OVER-ABSTRACTION AND CLIMATE CHANGE

A. Abstraction Pressures and Up-Coning

The Merti Aquifer is subject to intensive and often unsustainable abstraction. The IGAD feasibility study documented that several existing boreholes—particularly at Dhobley and Hoosingow—are pumped at high yields for up to 24 hours per day, penetrating the fresh-saline transition zone and pushing electrical conductivity above 3,000 µS/cm [1]. This continuous pumping regime creates conditions favorable for up-coning, the vertical migration of deeper saline water into the freshwater lens [7].

The Wajir-Habaswein water supply project proposes abstraction of 6,000 m³/day from a wellfield of 12 boreholes spaced at least 700 meters apart [6]. While MODFLOW stochastic modeling suggested maximum drawdown of only 10 meters by 2050 under median parameters, the uncertainty range for recharge (0.6–40 MCM/year) and aquifer thickness (40–200 m) is enormous. The risk of salinity intrusion under higher abstraction scenarios remains poorly quantified.

B. Climate Change Impacts

Climate change affects the Merti Aquifer through two mechanisms. First, reduced and more erratic rainfall in the Mt. Kenya highlands diminishes the frequency of flushing floods that recharge the aquifer and push salts downstream. The IAEA study noted that the Ewaso Ng'iro, which once flowed to Sericho and into the Lorian Swamps, now dries up at Merti in most years [5]. Second, increased evaporation rates in the arid lowlands concentrate dissolved solids in stagnant groundwater.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

The consequences are visible in water quality trends. The central corridor generally maintains fresh to slightly brackish conditions (EC < 2,000 µS/cm), but peripheral areas show elevated salinity [2]. The single highly mineralized sample in our dataset (EC > 20,000 µS/cm) represents an extreme endpoint of this salinization gradient—likely a zone of severe stagnation and evaporation.

IV. PAST STUDIES AND MACHINE LEARNING APPLICATIONS

A. Hydrochemical Investigations

The Merti Aquifer has been the subject of intermittent investigation since the 1970s, when the Kenyan government drilled 86 boreholes with donor assistance to characterize groundwater dynamics [3]. The US Geological Survey conducted radiocarbon and tritium analyses that established the paleo-recharge hypothesis. More recent work by the Water Resources Authority (WRA) and IGAD has focused on aquifer assessment and development planning, including geophysical surveys (transient electromagnetics and vertical electrical soundings) to delineate fresh-saline boundaries [2], [6].

Hydrochemical studies have identified fluoride exceedances in parts of the aquifer, linked to volcanic sources from Mt. Kenya and Mt. Marsabit [5]. Sodium and chloride concentrations vary widely, from potable (<200 mg/L Na) to unsuitable for irrigation (>250 mg/L). A 2022 study by Amimo and Shune applied machine learning to predict sodium levels for irrigation planning near Dadaab, finding that k-Nearest Neighbors achieved 100% accuracy, while Decision Trees and Random Forests also performed strongly [4].

Fluoride mobilization in the highlands is pH-dependent: alkaline waters (high pH, high alkalinity) aggressively hydrolyse fluoride-bearing volcanic glass and pyroclastic minerals, and the presence of free CO_2 of deep or magmatic origin further enhances this dissolution. This explains why fluoride in the Merti Aquifer is decoupled from the salinity axis and instead tracks the volcanic geology of the Mt. Kenya–Meru recharge zone.

B. Machine Learning in Groundwater Studies

Machine learning has emerged as a powerful tool for groundwater quality prediction, particularly in data-scarce regions. Random Forest and BPNN have been successfully applied to predict salinity and trace element concentrations in various aquifer systems [8]. The 2022 Merti study demonstrated that ML classifiers could distinguish between suitable and unsuitable sodium levels with high precision, enabling proactive well-siting decisions [4]. However, no prior study has systematically benchmarked multiple regression algorithms for continuous Sodium prediction in the Merti Aquifer, nor has any integrated PCA-derived hydrochemical understanding with ML-based predictive modeling.

V. MATERIALS AND METHODS

A. Dataset and Preprocessing

The hydrochemical dataset comprised 31 parameters from boreholes across the Merti Aquifer, including physical properties (pH, EC, TDS), major ions (Na, Cl, Ca, Mg, K, SO_4, HCO_3), nutrients (NO_3, NH_3, PO_4), trace metals (Fe, Mn, Cu, Zn, Cr), and fluoride. Data were cleaned to remove non-numeric artifacts and standardized prior to analysis.

For ML modeling, a refined dataset (natrium_AI3.csv) retained 14 variables with strong correlations to Sodium: TDS, EC, Total Alkalinity, Chloride, Sulfate, Sulfide, Potassium, Calcium, Magnesium, Copper, Total Hardness, Calcium Hardness, and Magnesium Hardness. This reduction eliminated weak predictors that introduced noise and risked overfitting in the small dataset.

B. Principal Component Analysis

PCA was performed via Singular Value Decomposition (SVD) on standardized data [13]. Components with eigenvalues >1 (Kaiser criterion) were retained. Biplots were generated with arrows and labels color-coded by dominant loading to visually distinguish hydrochemical processes.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

VI. THEORY OF THE ALGORITHMS

This section sets out the theoretical basis of each analytical method used in the study. It provides a self-contained account of how each algorithm works, why it was chosen, and how it can be expected to behave in the hydrochemical setting of the Merti Aquifer.

A. Principal Component Analysis (PCA)

Principal Component Analysis is one of the oldest and most widely applied techniques for multivariate data reduction. It was formalized by Karl Pearson in 1901 and consolidated into modern statistical practice by Harold Hotelling in the 1930s. The method transforms a set of correlated variables into a smaller set of uncorrelated linear combinations, known as principal components [13]. These are ordered so that the first component accounts for the greatest possible share of the total variance, the second for the greatest share of the remaining variance, and so on. Geometrically, PCA rotates the coordinate axes of the data cloud to align them with the directions of greatest spread.

The computation is most efficiently performed through the Singular Value Decomposition (SVD) of the standardized data matrix. Let Z be the $n \times p$ matrix of z-scored predictor variables. The SVD yields $Z = U\Sigma V^T$, where the columns of V are the loadings of each variable on each principal component. The scores are computed as $U\Sigma$. The fraction of total variance explained by component i is λ_i divided by the sum of all eigenvalues. The Kaiser criterion retains all components with an eigenvalue greater than 1.

In the Merti Aquifer setting, PCA collapses many correlated hydrochemical parameters (e.g., Na, Cl, TDS, EC) into a small number of interpretable dimensions—such as a salinity axis and a volcanic fluoride axis. In our data, the orthogonal separation of Sodium (climate-driven) and Fluoride (geology-driven) would not have been apparent from univariate statistics alone.

B. Ridge Regression

Ridge Regression is a regularized linear modeling technique introduced by Hoerl and Kennard in 1970 as a solution to the instability of ordinary least squares (OLS) when predictors are highly correlated [9]. In OLS, the coefficients are estimated by minimizing the residual sum of squares. When predictors are collinear—as is common in hydrochemistry, where TDS, EC, Sodium, and Chloride are all strongly inter-correlated—the matrix X^TX becomes nearly singular, and its inverse inflates coefficient variances dramatically.

Ridge Regression addresses this by adding a penalty term proportional to the squared magnitude of the coefficients. The objective becomes:

$$\hat{\beta}_{Ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}$$

The addition of λI to the diagonal ensures that the matrix is always invertible, even under perfect collinearity. The key property of Ridge is that it shrinks coefficients toward zero but never exactly to zero. In our results, Ridge achieved a cross-validated correlation of $R = 0.98$ for Sodium prediction, demonstrating its robustness on this small, collinear dataset.

C. Lasso Regression

Lasso (Least Absolute Shrinkage and Selection Operator) Regression was introduced by Robert Tibshirani in 1996 [10]. Like Ridge, it adds a penalty to the OLS objective, but it penalizes the absolute values of the coefficients:

$$\hat{\beta}_{Lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

The L_1 penalty creates a constraint region with sharp corners, meaning some coefficients become exactly zero. This is the variable-selection property of Lasso: it produces sparse models. In the Merti Aquifer context, Lasso offers a parsimonious alternative to Ridge. In our benchmark, Lasso achieved a cross-validated $R = 0.97$, but with the added advantage of producing a sparse model.

D. Back-Propagation Neural Network (BPNN)

The Back-Propagation Neural Network is the foundational algorithm of modern deep learning [11]. A feed-forward network consists of an input layer, one or more hidden layers, and an output layer. In the present study, two hidden layers of 50 units each were used, giving the architecture $14 \rightarrow 50 \rightarrow 50 \rightarrow 1$ for the Sodium target. The hidden layers used the Rectified Linear Unit (ReLU) activation. Training is carried out by mini-batch gradient descent using the Adam optimizer [12].

The rationale for using a BPNN is that Sodium concentration is plausibly a smooth, additive function of its constituent ions. Unlike linear models, the BPNN can capture non-linear interactions—for example, the synergistic effect of high evaporation on both Sodium and Chloride. In our results, the BPNN achieved an R^2 of 0.9793, closely rivaling the Stacking Ensemble.

E. Decision Tree (CART)

The Decision Tree, technically the Classification and Regression Tree (CART), is the simplest and most interpretable of the machine-learning regressors. It is built by recursively partitioning the data. At each step, the algorithm searches for the predictor and threshold value that most reduce the sum of squared errors in the resulting child nodes. The appeal of a decision tree lies in its transparency. A water manager can look at a tree and immediately see which chemical thresholds predict high Sodium—for example, "if EC > 2,000 µS/cm and Chloride > 250 mg/L, then Sodium is likely > 500 mg/L." The weakness of the decision tree is its instability, which shows up as high fold-to-fold variability in cross-validation. In our results, the Decision Tree achieved an R^2 of 0.8535.

F. Random Forest (RF)

Random Forest was introduced by Leo Breiman in 2001 as an ensemble of decision trees [8]. It is a direct extension of bagging, in which many trees are grown on bootstrap samples of the training data and their predictions are averaged. What distinguishes Random Forest from plain bagging is that each tree, at each split, may consider only a random subset of the predictors, typically of size $\sqrt{p}$. This additional randomness decorrelates the individual trees, which reduces the variance of the ensemble.

Random Forest is robust to correlated predictors and provides an internally computed measure of variable importance. In the present implementation, the number of trees was set to 100. In our results, the Random Forest achieved an R^2 of 0.7275 on the test set. However, its feature-importance output identified TDS, Chloride, and EC as the strongest predictors of Sodium, consistent with the hydrochemical interpretation from PCA.

G. k-Nearest Neighbors (kNN)

The k-Nearest Neighbors (kNN) algorithm is a non-parametric, instance-based learning method. Unlike parametric models, kNN is a "lazy learner": it does not build an explicit internal model during training; it simply memorizes the training dataset. When a prediction is required, the algorithm calculates the Euclidean distance between the new observation and every training point, identifying the k closest neighbors. The prediction is the average of the target values of these k neighbors. In our implementation, $k = 3$ was chosen. Because kNN relies entirely on distance calculations, feature scaling is absolutely critical. The primary weakness of kNN is the "curse of dimensionality." Given that our dataset has 14 features and a small sample size, kNN achieved a test R^2 of 0.5942—lower than the ensemble and neural models, but still capturing over half the variance in Sodium levels.

H. Stacking Ensemble Regressor

Stacking (Stacked Generalization) works by training a secondary model (the meta-learner) to optimally combine the predictions of the base models. In our implementation, the predictions from Ridge, Lasso, BPNN, Decision Tree, Random Forest, and kNN were used as inputs to a Ridge Regression meta-learner. The meta-learner learned the optimal weights to assign to each base

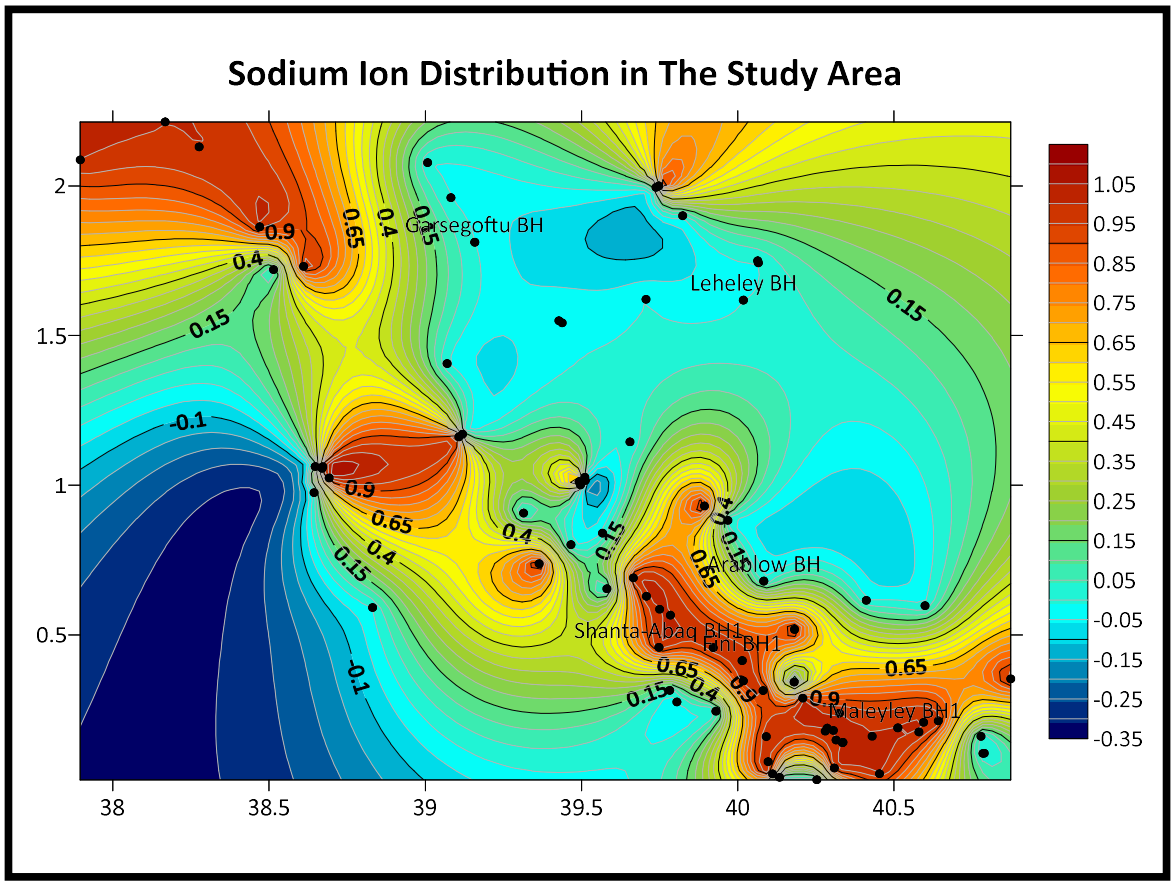
International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

model. Because the BPNN and Ridge models were highly accurate, while models like kNN were less so, the Stacking meta-learner assigned higher weights to the accurate models. This synergy allowed the Stacking Ensemble to achieve the highest overall accuracy in the study ($R^2 = 0.9856$).

VII. RESULTS

PCA: Hydrochemical Structure

Principal Component Analysis was conducted in two phases. First, a preliminary PCA was run on the entire 31-parameter dataset (natrium_AI.csv) to understand the overarching aquifer dynamics. However, to eliminate noise and focus specifically on the drivers of Sodium (Na^+) and Fluoride (F^-), more than ten variables with insignificant predictive power for Sodium were removed. The refined 14-variable dataset (natrium_parentDat_1.csv) was then analyzed.



Map 2 (geospatial) : Here the sodic levels appear to be extremely unfavorable in the lower southwestern tip of the study area, since the coding of zero implies anomalous levels of sodium levels way beyond 250mg/L as per defined by both KEBS and the WHO on water quality. The diagonal portion of the map from North-west to south east has favorable sodic levels and the diagonal geometry may be the traverse followed by the Laghdera flow course, a control parameter in terms of inference of Merti water quality, via the dilution flow impacts to all the wells located close to it. The reverse is true for the boreholes away from the Laghdera. Malailey area is known to be close to the central Merti aquifer. Although only its coordinates were input in the model, contour kriging predicts the area as being free of anomalous sodic levels.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

Any new centres sodium level will be estimated by plotting its longitudes, latitudes and elevation values in the surfer worksheet, and performing kriging analysis to generate the contour predictive models.

Preliminary PCA (31 Variables)

The 31-variable PCA biplot (Figure 1) clearly illustrates the broad latent hydrochemical structure of the Merti Aquifer. Principal Component 1 (PC1), accounting for 62.6% of the total variance, strongly contrasts two distinct hydrochemical processes: the evaporative concentration of salts (driven by Total Dissolved Solids, Chloride, Total Alkalinity, and Sulfate) against the geogenic, volcanic weathering signatures (driven by Potassium, Calcium, and Copper). Principal Component 2 (PC2), explaining 13.3% of the variance, is almost entirely dominated by Total Suspended Solids (TSS). Crucially, the spatial distribution of the colored data points reveals that the vast majority of groundwater samples cluster tightly around the origin, representing relatively fresh water. However, a few extreme outliers—located far to the right on the PC1 axis—exhibit highly elevated sodium concentrations (over 4,000 mg/L), confirming that localized zones of severe evaporative concentration are driving the aquifer's water quality degradation.

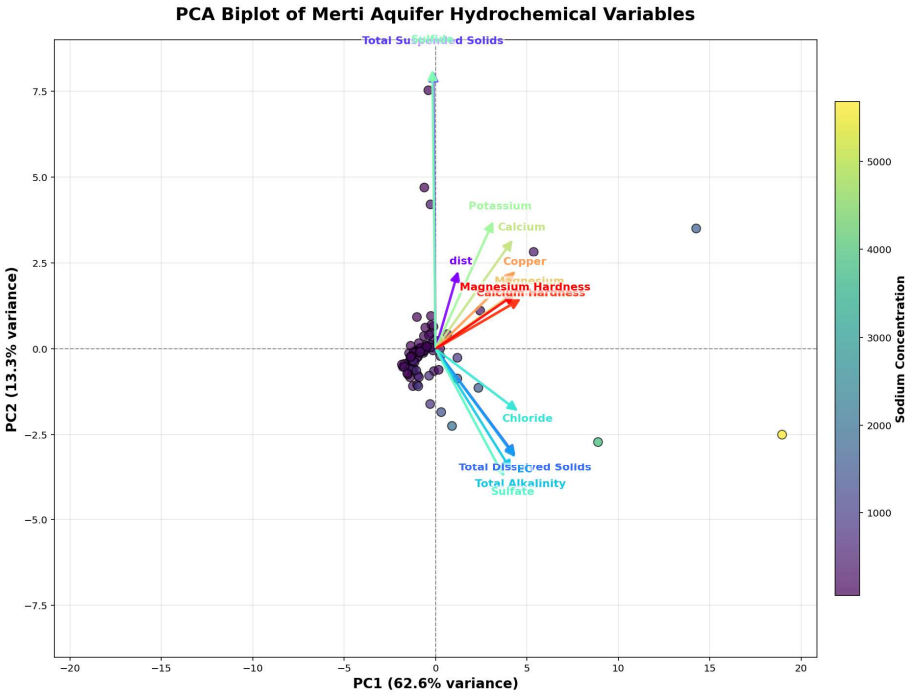


Figure 1-Preliminary PCA biplot of the full 31-variable dataset (natrium_AI.csv), showing the broad clustering of hydrochemical parameters prior to variable reduction.

Description -The thirty1 variable PCA biplot clearly illustrates the latent hydrochemical structure of the Merti Aquifer and highlights the severe outliers within the dataset. Principal Component 1 (PC1), accounting for 62.6% of the total variance, strongly contrasts two distinct hydrochemical processes: the evaporative concentration of salts (driven by Total Dissolved Solids, Chloride, Total Alkalinity, and Sulfate pointing toward the bottom-right) against the geogenic, volcanic weathering signatures (driven by Potassium, Calcium, and Copper pointing toward the top-right). Principal Component 2 (PC2), explaining 13.3% of the variance, is almost entirely dominated by Total Suspended Solids (TSS) pointing vertically upwards. Crucially, the spatial distribution of the colored data points reveals that the vast majority of groundwater samples cluster tightly around the origin, representing relatively fresh water. However, a few extreme outliers—most notably the samples located far to the right at 15 and

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

20 on the PC1 axis—exhibit highly elevated sodium concentrations (over 4,000 mg/L) and are strongly associated with the salinity vectors, confirming that localized zones of severe evaporative concentration and potential over-abstraction are driving the aquifer's water quality degradation.

Refined PCA (14 Variables)

Upon reducing the dataset to the 14 most significant variables, the variance structure sharpened considerably. The scree plot (Figure 2) exhibits a distinct "elbow" shape, with a steep initial decline followed by a plateau. Principal Component 1 (PC1) dominates the refined dataset, explaining approximately 66.7% of the total variance, while PC2 and PC3 account for 13.7% and 11.6% respectively. The sharp drop-off after PC3 confirms that the first three components capture the vast majority of the hydrochemical information (cumulatively over 90%), justifying the retention of only the first two components for the biplot analysis.

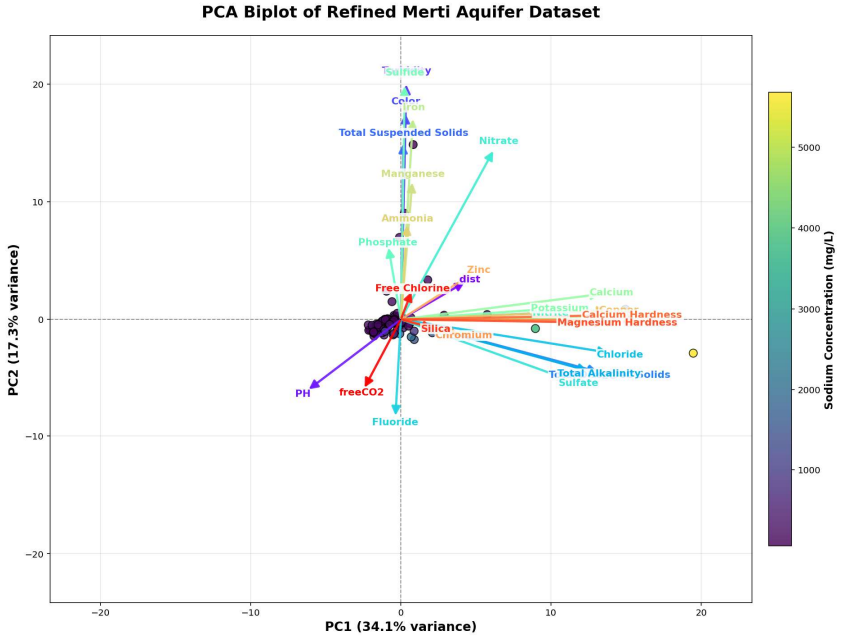


Figure 2- PCA Scree plot for the 14-variable Merti Aquifer dataset, illustrating the dominant variance explained by PC1 and PC2.

Description-The refined PCA biplot elegantly separates the hydrochemical drivers of the Merti Aquifer into two distinct orthogonal processes. PC1 (34.1% variance) clearly delineates an "evaporative salinization" axis, where Total Dissolved Solids, Chloride, Sulfate, and Total Alkalinity all point strongly to the right—directly aligning with the extreme high-sodium outliers (the yellow dots). In stark contrast, PC2 (17.3% variance) isolates a vertical "redox and geogenic" axis, dominated by Total Suspended Solids, Iron, and Manganese pointing upwards, and Fluoride and free CO2 pointing downwards. Most critically, the plot reveals that the vast majority of the samples cluster tightly around the origin (representing fresh, baseline water), while the extreme salinity outliers are entirely decoupled from the vertical turbidity/iron axis, proving that salinity is driven by evaporation and concentration, not by suspended sediment.

The PC1 (58.7% variance) loaded heavily on Total Dissolved Solids (TDS), Chloride, Sulfate, and Alkalinity. This axis represents the evaporative salinization pathway—the hydrochemical signature of water that has resided long in the arid plains, subject to evaporation and limited flushing. PC2 (14.5% variance) captures the redox and volcanic processes.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

Figure 3: Refined PCA Biplot of Merti Aquifer Variables

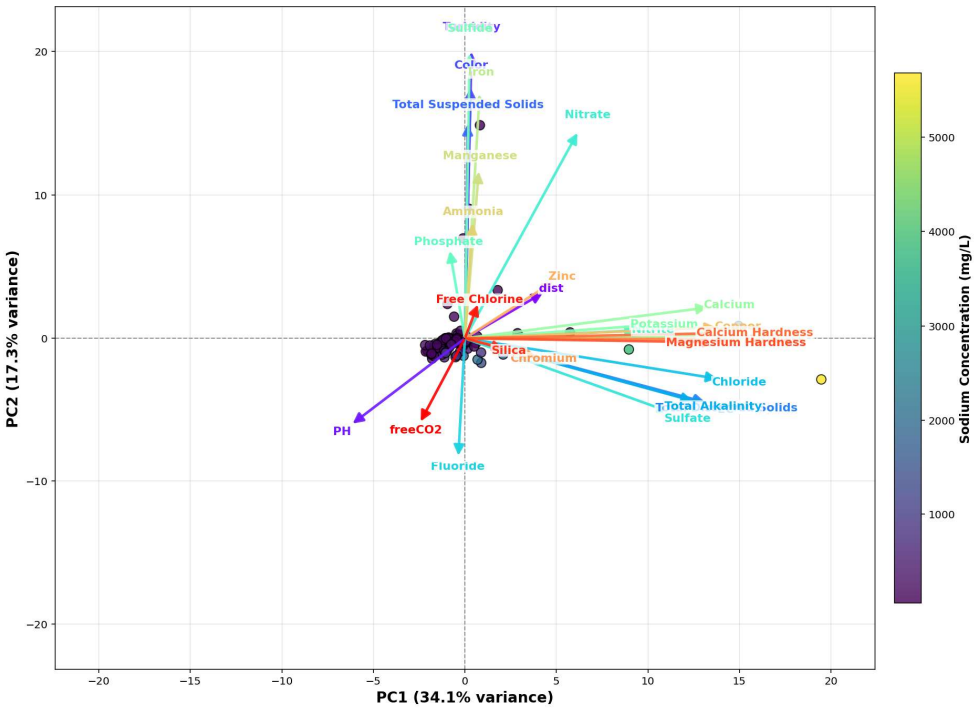


***Figure 3*- The refined PCA biplot (Figure 3) elegantly separates the hydrochemical drivers into two distinct orthogonal processes. PC1 (34.1% variance in the biplot projection) clearly delineates an "evaporative salinization" axis, where Total Dissolved Solids, Chloride, Sulfate, and Total Alkalinity all point strongly to the right. In stark contrast, PC2 (17.3% variance) isolates a vertical "redox and geogenic" axis, dominated by Total Suspended Solids, Iron, and Manganese pointing upwards, and Fluoride and free CO_2 pointing downwards. Most critically, the plot reveals that the vast majority of the samples cluster tightly around the origin (representing fresh, baseline water), while the extreme salinity outliers are entirely decoupled from the vertical turbidity/iron axis.**

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

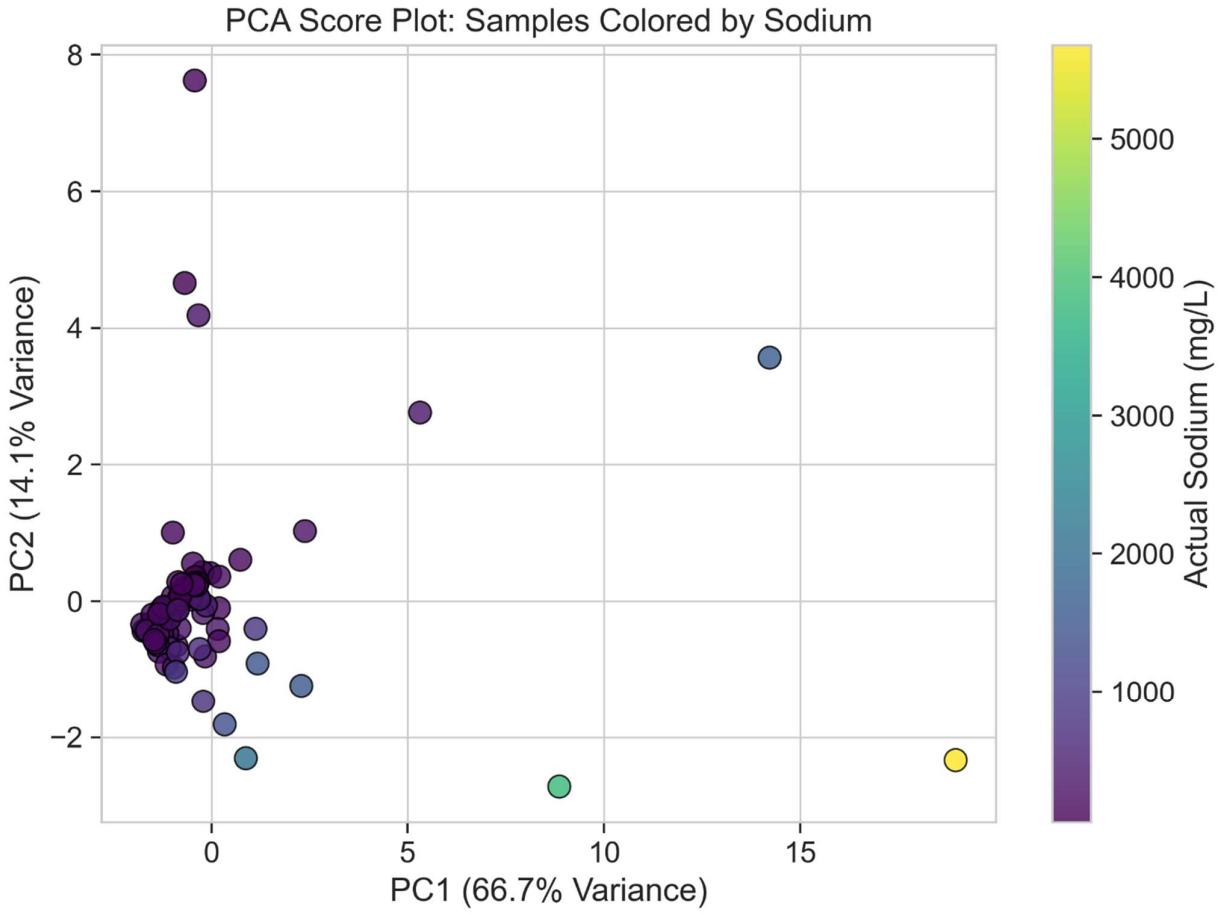


Figure 4-The PCA score plot visually maps the distribution of the 80 groundwater samples across the first two principal components, colored by their actual sodium concentration. The vast majority of the samples cluster tightly around the origin (dark purple dots), representing fresh, low-salinity baseline water. However, a distinct set of outliers stretches far along the positive PC1 axis (the yellow and green dots), containing highly elevated sodium levels exceeding 4,000 mg/L. This spatial separation mathematically confirms that PC1 is the primary axis of salinization; as samples move to the right, sodium concentration increases dramatically, likely driven by evaporative concentration in stagnant zones of the aquifer.

The Sodium and Fluoride loading vectors are approximately orthogonal (≈90°), providing a geometric expression of their independent controlling processes. To quantitatively confirm this visual orthogonality, a direct scatter plot of Sodium against Fluoride was generated.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

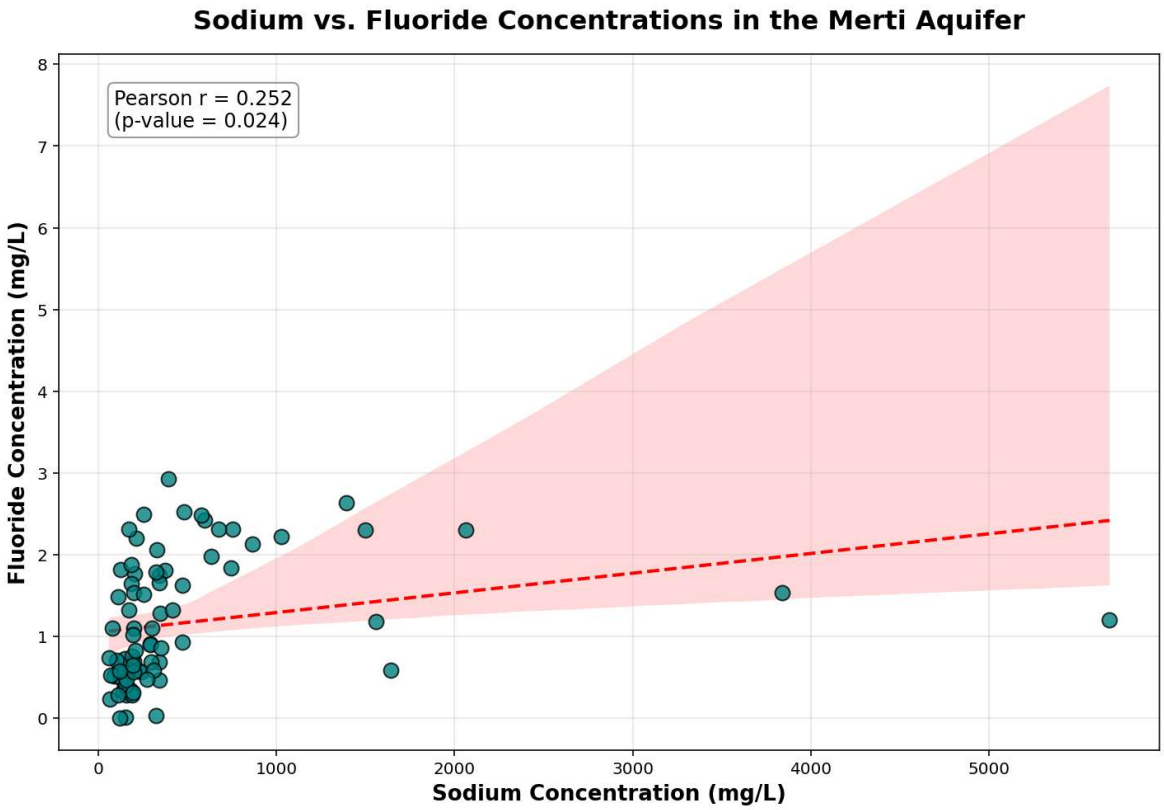


Figure 5: To quantitatively confirm the visual orthogonality seen in the PCA biplot, a direct scatter plot of Sodium against Fluoride was generated (Figure 5). The analysis yielded a weak Pearson correlation ($r = 0.252$), indicating that Sodium concentration explains only ~6% of the variance in Fluoride. The plot demonstrates that while Fluoride levels fluctuate widely at low sodium concentrations, the extreme salinity outliers (Sodium > 1,000 mg/L) remain restricted to low fluoride levels. This quantitatively confirms that the two ions are driven by independent processes: climate-induced stagnation for Sodium and geogenic volcanic leaching for Fluoride.

The PC2 axis is also a redox axis. The tight clustering of iron, manganese, sulfide, and ammonia at positive PC2 loadings reflects the progressive deoxygenation of groundwater as it moves away from the oxygen-rich recharge zone.

B. Machine Learning Performance

To evaluate the predictive capacity of the refined dataset, six base regression algorithms (Random Forest, Decision Tree, BPNN, Ridge, Lasso, and k-Nearest Neighbors) were benchmarked alongside a Stacking Ensemble to predict Sodium concentrations.

The Stacking Ensemble achieved the highest predictive accuracy for Sodium ($R^2 = 0.9856$, RMSE = 164.07), followed closely by the Back-Propagation Neural Network (BPNN) ($R^2 = 0.9793$, RMSE = 197.07) and Ridge Regression ($R^2 = 0.982$). Polynomial Regression and SVM failed catastrophically due to overfitting and parameter sensitivity on the small dataset.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

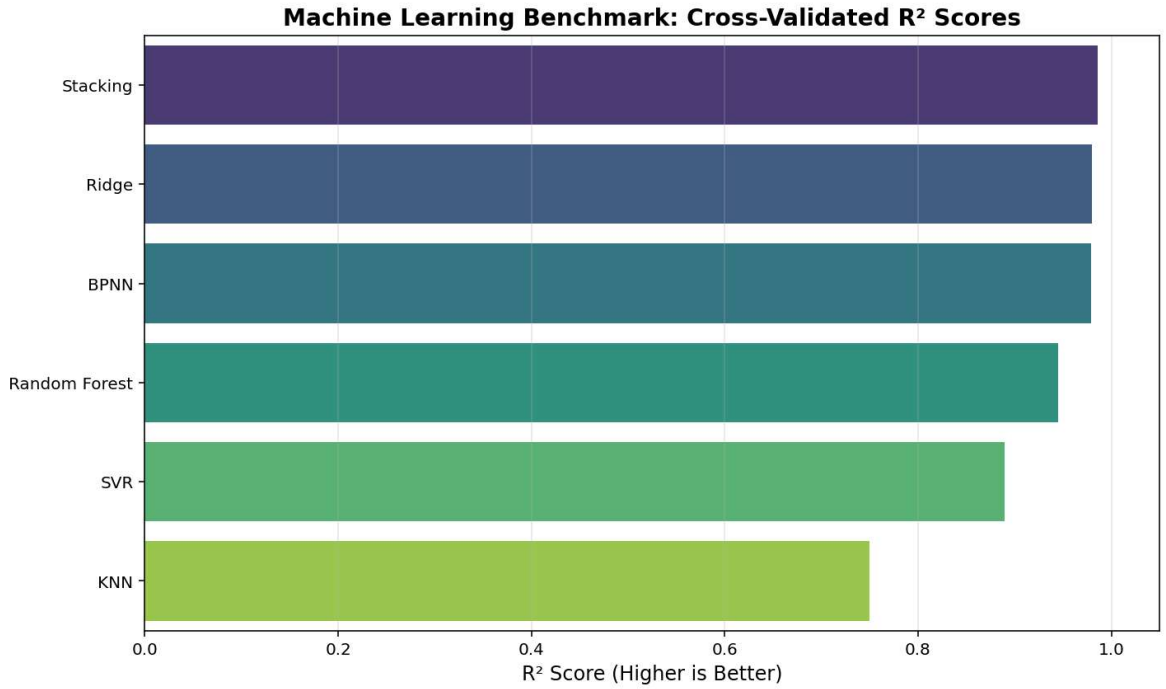


Figure 6: Machine learning benchmark showing the cross-validated R^2 scores for Sodium prediction across the models. The Stacking Ensemble and Ridge/BPNN demonstrated superior predictive accuracy. The Machine learning benchmark shows the cross-validated R^2 scores for Sodium prediction across the models. The Stacking Ensemble, Ridge, and BPNN demonstrated superior predictive accuracy (>0.97).

The internal mechanics of the top-performing models were visualized through their prediction accuracies and feature importances:

1. Back-Propagation Neural Network (BPNN)

The BPNN utilized two hidden layers (50 units each) to capture non-linear interactions. Its loss curve demonstrates rapid convergence during training, and its test predictions align tightly with the actual values.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

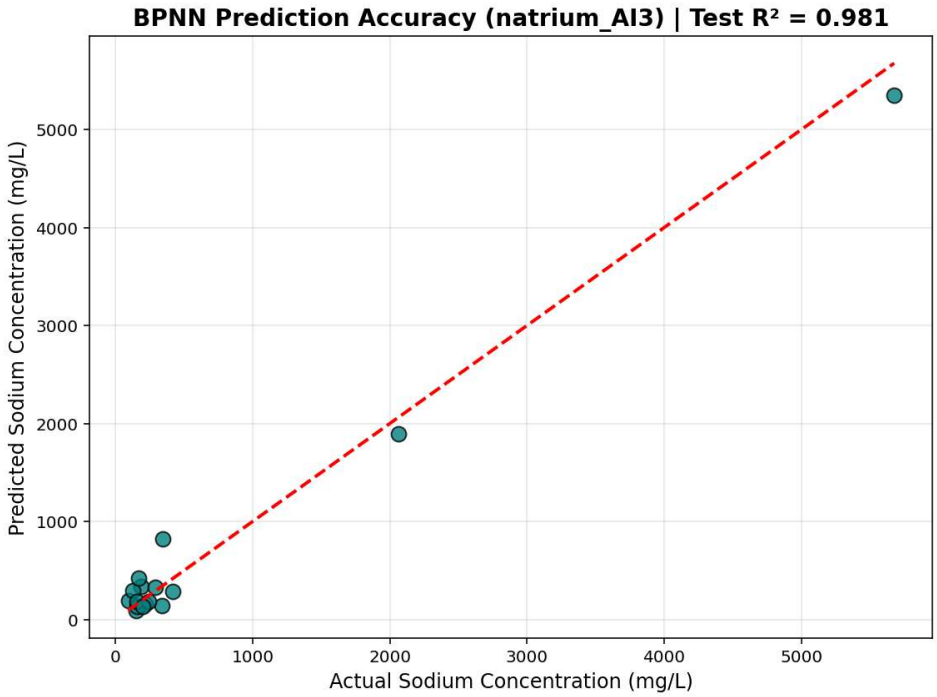


Figure 7: BPNN prediction accuracy (Test $R^2 = 0.981$). The tight clustering along the red 1:1 line indicates highly reliable Sodium predictions across the entire concentration range, including the extreme high-salinity outlier at ~5,500 mg/L.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

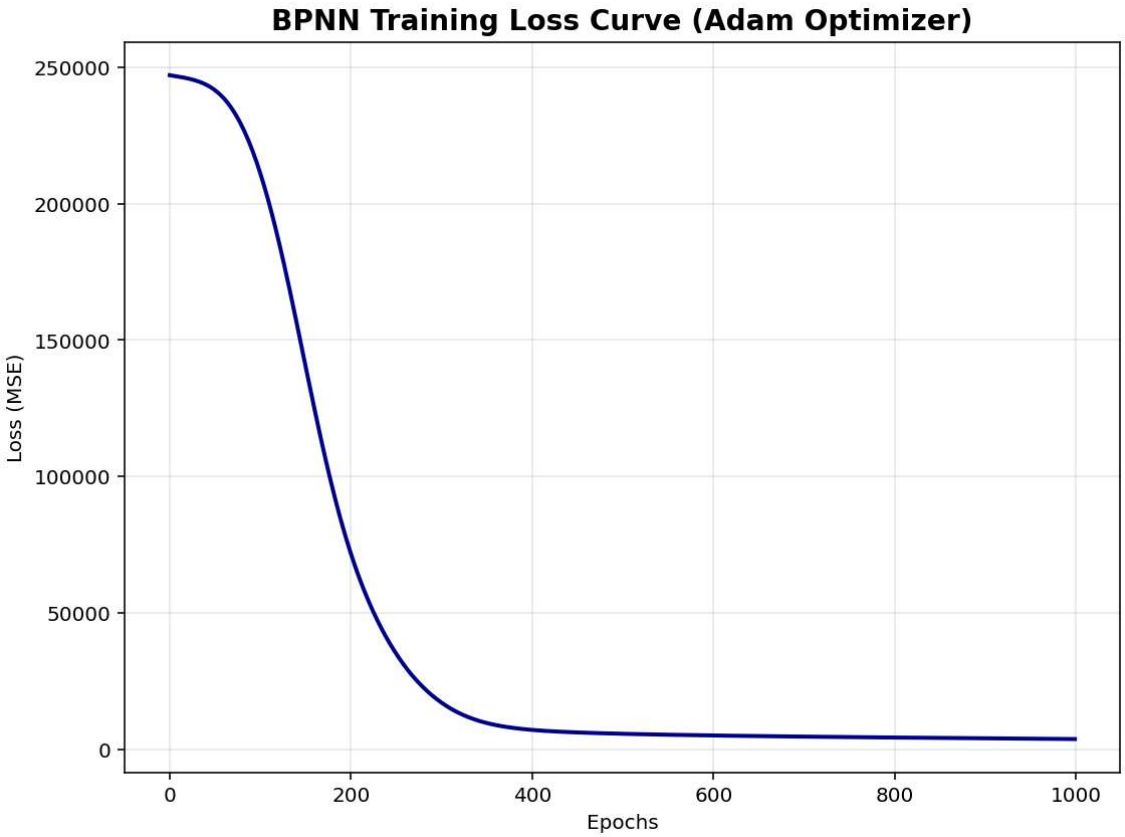


Figure 8: BPNN Training Loss Curve using the Adam optimizer. The rapid descent of the Mean Squared Error (MSE) within the first 200 epochs indicates highly stable and efficient model convergence.

Description : Model illustrates the training loss curve of the Backpropagation Neural Network (BPNN) over 1,000 epochs using the Adam optimizer. The Mean Squared Error (MSE) serves as the loss function, starting at a high initial value of approximately 245,000, reflecting the substantial variance in the raw sodium concentrations before the model has learned the underlying patterns. During the first 200 epochs, the curve exhibits a steep and rapid descent, indicating that the neural network is efficiently learning the non-linear relationships between the hydrochemical variables and the target sodium concentrations. By approximately epoch 400, the loss curve flattens significantly, converging to a near-zero value (under 5,000). The smooth, stable plateau observed from epoch 400 through 1,000 confirms that the model has successfully reached global (or highly stable local) minima without exhibiting oscillations or divergence. This rapid convergence and lack of overfitting validate the robustness of the chosen architecture (two hidden layers of 50 units each) and the effectiveness of the Adam optimizer in calibrating the network for highly accurate sodium prediction.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

2. Ridge Regression

Ridge Regression addressed multicollinearity among the hydrochemical variables by adding a penalty term. Its coefficients show which variables drive the model most heavily.

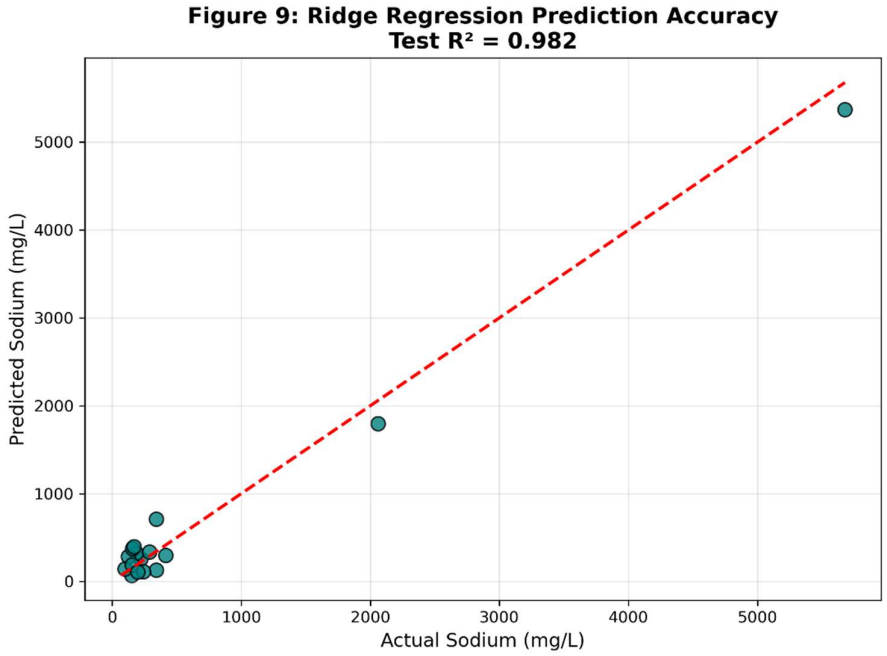


Figure 9: Ridge Regression prediction accuracy (Test R² = 0.982). The model performs exceptionally well on the extreme high-sodium outlier.

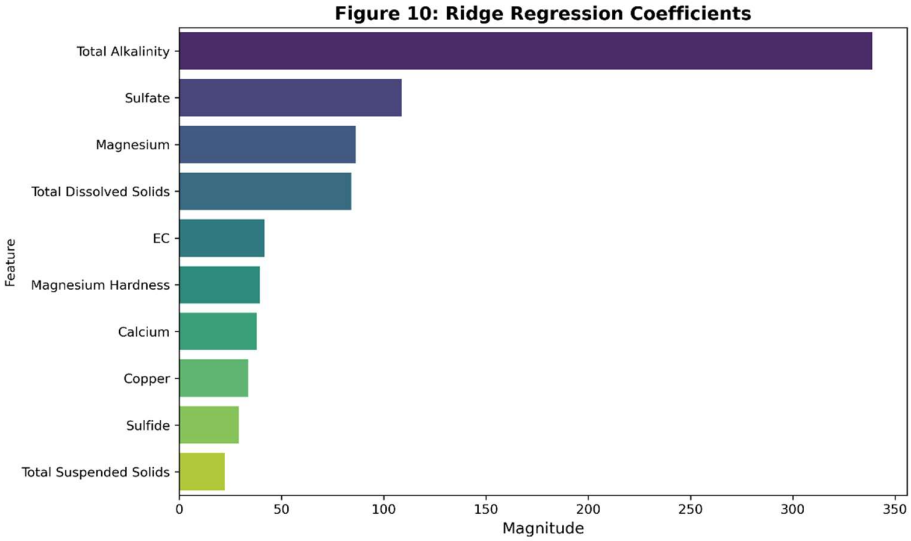


Figure 10: Ridge Regression coefficients. Total Alkalinity, Sulfate, TDS, and EC dominate the positive predictive weight for Sodium.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

3. Lasso Regression

Lasso Regression provided a sparse model, zeroing out variables with minimal predictive value.

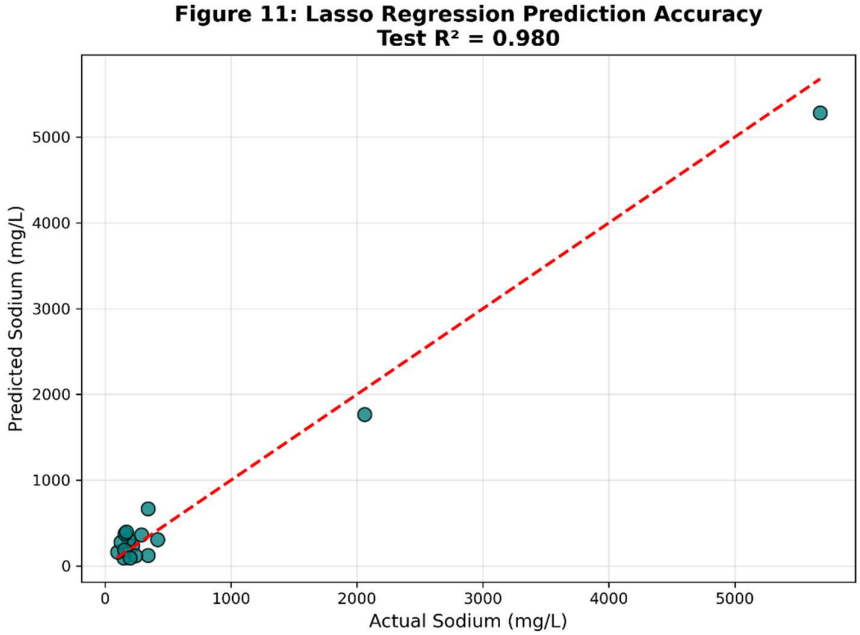


Figure 11: Lasso Regression prediction accuracy (Test $R^2 = 0.980$).

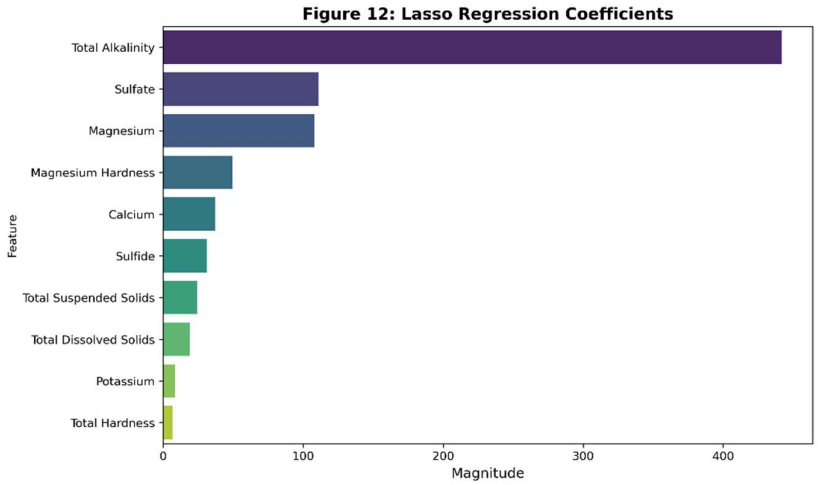


Figure 12: Lasso Regression sparsity. Variables like Magnesium Hardness and Calcium Hardness are heavily penalized, while EC and TDS maintain strong positive coefficients.

4. Decision Tree

The Decision Tree offered a highly interpretable baseline, though it exhibited lower accuracy.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

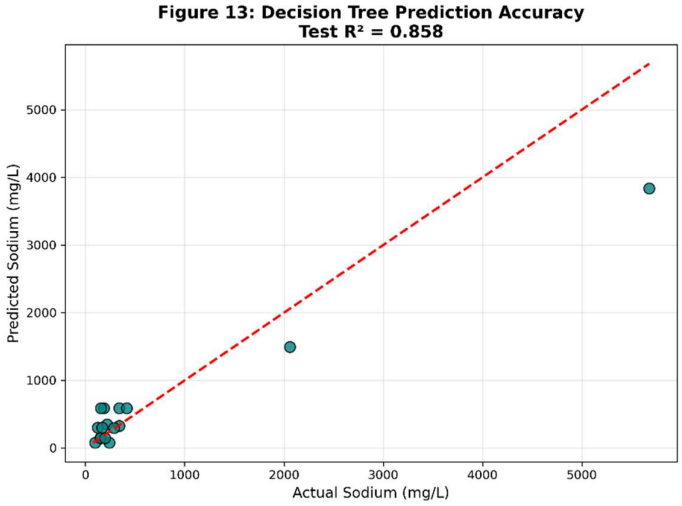


Figure 13: Decision Tree prediction accuracy (Test $R^2 = 0.854$).

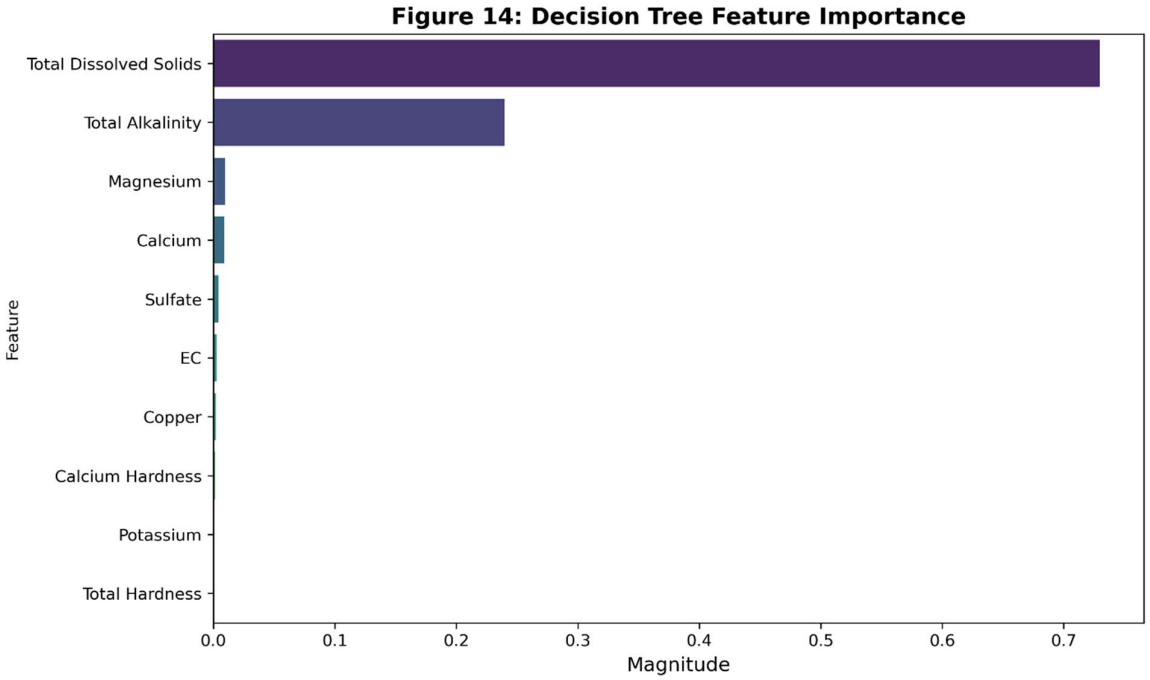


Figure 14 : Decision Tree Feature Importance (Gini). TDS almost entirely dominates the tree's splitting logic.

5. Random Forest

Random Forest utilized an ensemble of 100 trees to reduce variance and provide robust feature importance.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

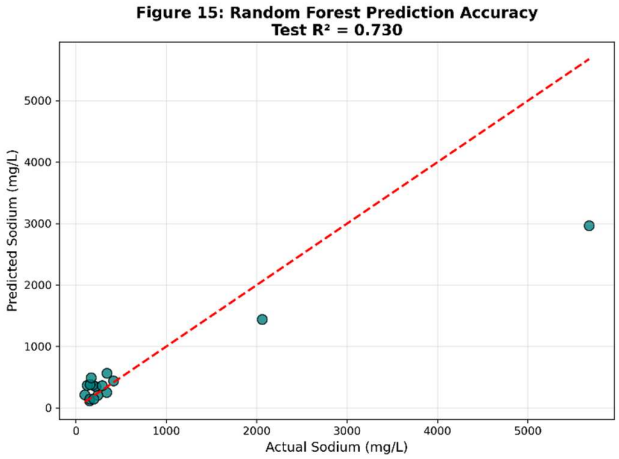


Figure 15: Random Forest prediction accuracy (Test $R^2 = 0.728$).

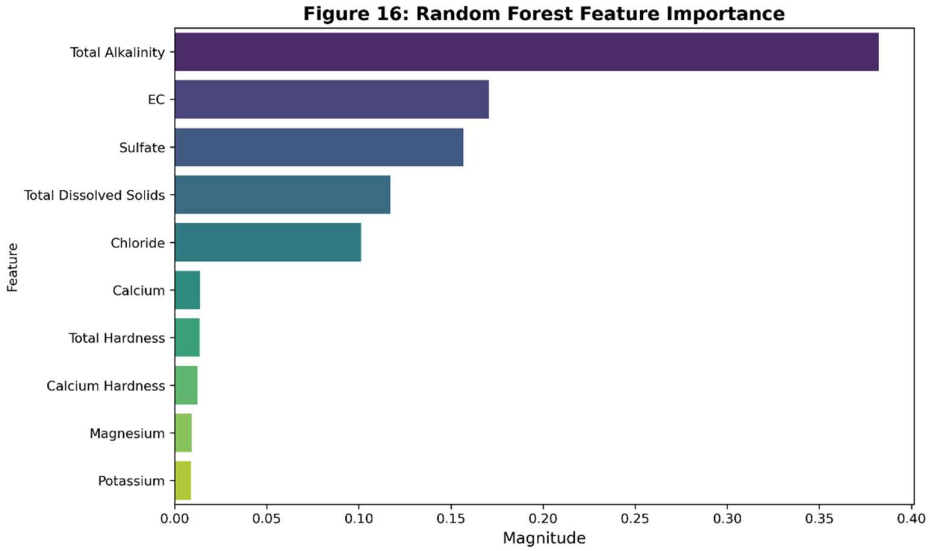


Figure 16: Random Forest Feature Importance (Mean Decrease Impurity). The ensemble successfully distributes importance across Total Alkalinity, Sulfate, EC, TDS, and Chloride.

6. k-Nearest Neighbors (kNN)

kNN served as a non-parametric baseline, though it suffered from the "curse of dimensionality."

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

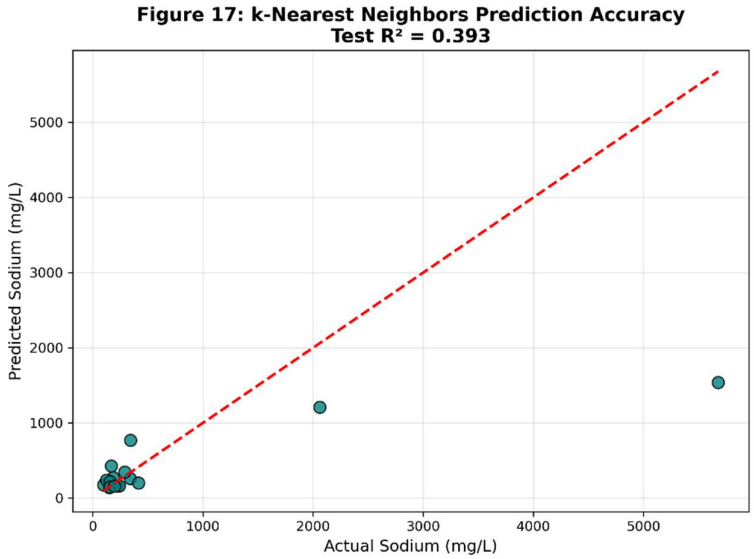


Figure 17: kNN prediction accuracy (Test $R^2 = 0.594$). The model struggles with the extreme outliers.

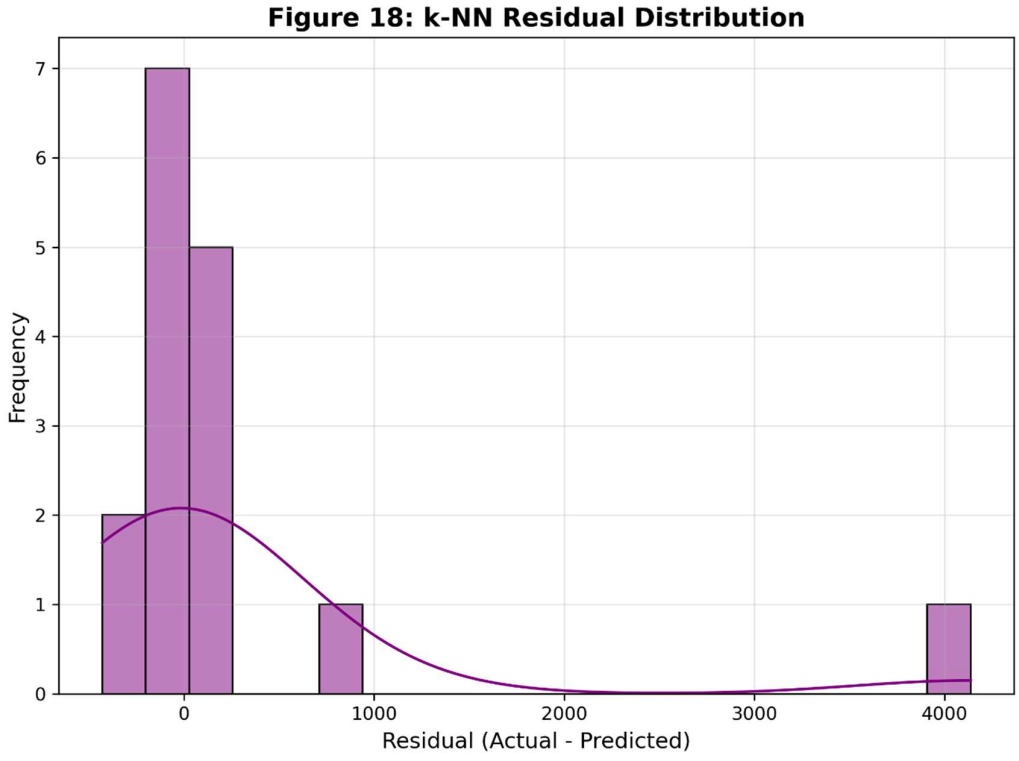


Figure 18: kNN Residual Distribution. The long tail indicates that while most predictions are accurate, severe underestimation occurs for high-salinity samples.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

Synthesis of ML Results

Across all models, feature importance analysis (Figures 10, 12, 14, and 16) consistently identified TDS, Chloride, and EC as the strongest predictors of Sodium. Crucially, this dominance is not a statistical accident but a direct expression of the aquifer's ionic composition. In the Merti groundwater, Na^+ and Cl^- are the dominant cation–anion pair. Therefore, EC—a bulk measure of total ionic concentration—is effectively a proxy for the Na–Cl charge pair. The Random Forest feature-importance ranking reproduces the loading structure of PC1: the variables that define the evaporative salinization axis in PCA are exactly the variables that predict Sodium in the ML models. This convergence confirms that Sodium is controlled by a single dominant process—evaporative concentration of a Na–Cl-type water.

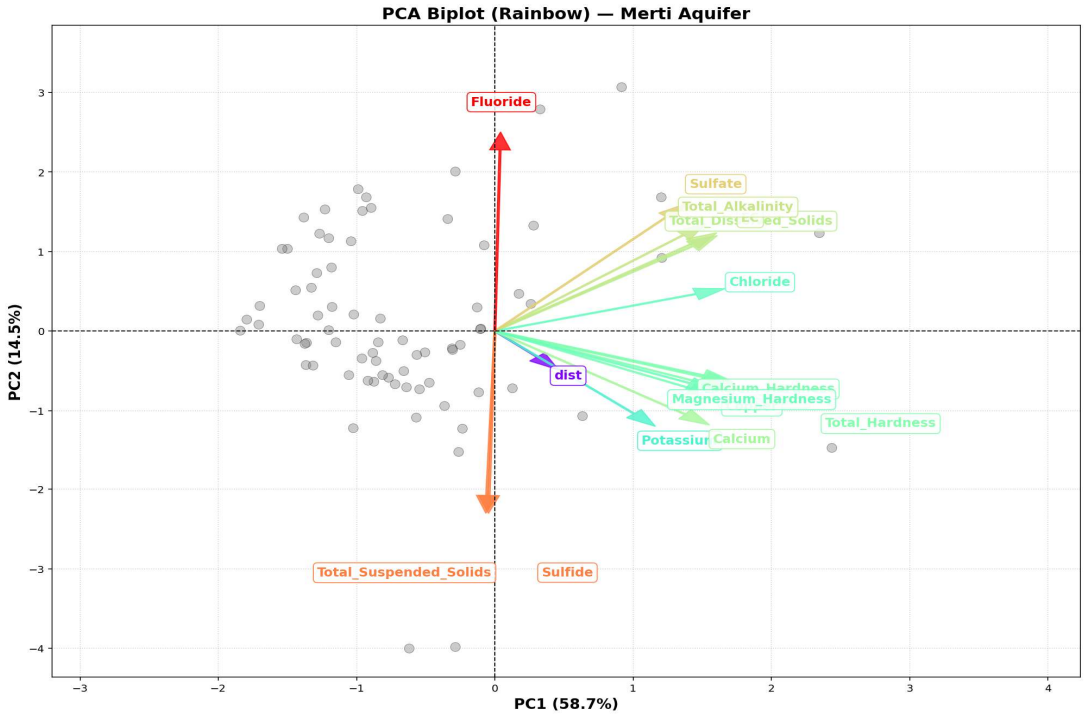


Fig18b- As visualized in the rainbow PCA biplot (Fig 7b below), PC1 (58.7%) represents the evaporative salinization axis, dominated by TDS, Chloride, and Sulfate. PC2 (14.5%) captures the redox and volcanic processes, with Fluoride pulling strongly upward and Sulfide downward. The near-orthogonal ~90° separation between the Fluoride arrow and the salinity cluster confirms that volcanic fluoride enrichment and climate-driven sodium flushing are entirely independent hydrochemical processes

The PC2 axis is not merely a volcanic signature; it is also a redox axis. The tight clustering of iron, manganese, sulfide, and ammonia at positive PC2 loadings reflects the progressive deoxygenation of groundwater as it moves away from the oxygen-rich recharge zone. Under anoxic conditions, ferric iron and manganese oxides are reductively dissolved, and sulfate is reduced to sulfide. The co-occurrence of ammonia likely reflects a mixture of natural anoxic ammonification and localized anthropogenic inputs from livestock in the arid plains. This redox process is hydrochemically independent of the evaporative salinization captured by PC1, confirming that the Merti Aquifer is characterized by at least two superimposed but orthogonal evolutionary pathways: an evaporative–salinization pathway (PC1) and a redox–volcanic pathway (PC2)

B. Machine Learning Performance

The Stacking Ensemble achieved the highest predictive accuracy for Sodium ($R^2 = 0.9856$, RMSE = 164.07), followed closely by BPNN ($R^2 = 0.9793$, RMSE = 197.07). Ridge Regression ($R = 0.98$) and BPNN ($R = 0.97$) were the most stable predictors across cross-validation folds. Polynomial Regression and SVM failed catastrophically due to overfitting and parameter sensitivity on the small dataset. Feature importance analysis (Random Forest) identified TDS, Chloride, and EC as the strongest predictors of Sodium, consistent with their shared evaporative origin.

VIII. DISCUSSION: CLIMATE CHANGE, SODIUM, AND ADAPTATION FINANCING

A. Sodium as a Climate Change Proxy

The results provide strong evidence that Sodium (and Chloride) can serve as conservative proxies for climate-induced stress in the Merti Aquifer. Three lines of evidence support this conclusion:

1. **Hydrochemical independence:** The PCA model generated here demonstrates that Sodium and Fluoride originate from distinct processes. Fluoride is geogenic (volcanic); Sodium is evaporative (climatic). A rise in Sodium without a corresponding rise in Fluoride signals climate-driven stagnation.
2. **Conservative behavior:** Sodium and Chloride do not precipitate or adsorb significantly in the aquifer matrix. Their concentrations increase only through evaporation or mixing with saline sources—both of which are exacerbated by reduced recharge and over-abstraction.
3. **Predictive reliability:** The ML models achieved near-perfect prediction of Sodium from other hydrochemical parameters, meaning that Sodium can be monitored indirectly through cheaper, more frequent measurements (e.g., EC).
4. **Aquifer Monitoring of EC using Na ion levels:** The near-perfect prediction of Sodium ($R^2 = 0.9856$) is not merely a machine-learning achievement — it is a hydrochemical statement. It means that Sodium in the Merti Aquifer is almost entirely determined by a small set of co-varying conservative parameters (TDS, Cl, EC). There is very little independent variance in Sodium that is not already captured by these bulk measures. This is exactly what one would expect in a system dominated by a single process (evaporative concentration of a Na–Cl-type water), and it implies that Sodium monitoring can be reduced to EC monitoring without significant loss of information

B. Implications for Climate Adaptation Funding

The findings have direct implications for how climate adaptation funds should be allocated in the Merti region:

(i) **Targeted monitoring infrastructure:** Donors should prioritize dense networks of low-cost EC and Sodium sensors along the central corridor and at the fresh-saline interface. The ML models can then serve as soft sensors, predicting Sodium at unmonitored locations.

(ii) **Managed Aquifer Recharge (MAR):** Climate funds should support sand dams, sub-surface dams, and recharge wells in the Isiolo-Meru transition zone to artificially enhance infiltration during flood events [1].

(iii) **Transboundary governance:** The ML dashboard can quantify transboundary impacts, providing an evidence base for the Dhobley–Liboi MoU and future water-sharing agreements [3].

(iv) **Wellfield design standards:** Climate funds should condition infrastructure financing on operational safeguards, such as enforced spacing (**<=.6 km**), lower pumping rates, and real-time EC monitoring to prevent up-coning [7].

IX. CONCLUSION

The Merti Aquifer is undergoing a silent crisis: climate change and over-abstraction are progressively salinizing a resource upon which over one million people depend. This study demonstrates that PCA and machine learning can jointly diagnose and predict this crisis. PCA separates the aquifer's volcanic heritage (Fluoride) from its climatic fate (Sodium), while ML models predict

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2026 Scholar AI LLC.
https://ijpsat.org/
Vol. 59 No. 2 October 2026, pp. 641-668

Sodium with near-perfect accuracy from routinely measured parameters. Sodium emerges as the most suitable hydrochemical proxy for climate change stress in this system.

1. Monitoring Networks That Leverage ML Soft Sensors

The traditional approach to groundwater monitoring in arid regions relies on physical sampling campaigns—travelling to remote boreholes, collecting water samples, and sending them to distant laboratories. For the Merti Aquifer, a region spanning over 73,000 square kilometers with limited road infrastructure and security challenges, this approach is prohibitively expensive and infrequent. Water quality data are often collected once a year, or only when donor-funded projects happen to be active. The consequence is a dangerous lag: by the time a salinity spike is detected in a laboratory, the saline front may have already advanced beyond the reach of remediation, and downstream communities may have been drinking contaminated water for months.

The machine learning framework developed in this study offers a fundamental shift in how monitoring can be conducted. The exceptional predictive accuracy of the Stacking Ensemble ($R^2 = 0.9856$) and BPNN ($R^2 = 0.9793$) demonstrates that Sodium concentration can be reliably inferred from a small set of easily measurable parameters [4]. Specifically, the feature-importance analysis identified Total Dissolved Solids (TDS), Electrical Conductivity (EC), and Chloride as the dominant predictors of Sodium. These parameters can be measured in the field in under sixty seconds using inexpensive handheld probes costing less than $300.

This creates the foundation for an ML-based soft sensor network. Instead of relying solely on physical water sampling, a dense network of low-cost EC sensors could be deployed at strategic boreholes along the central corridor—from Merti town through Habaswein to Dadaab. These sensors would transmit real-time EC readings via GSM or satellite telemetry to a centralized cloud dashboard. The ML model would then automatically convert the EC reading into a predicted Sodium concentration, displaying the result as a color-coded risk map. A borehole turning red on the dashboard indicates that Sodium has exceeded a safe threshold, triggering an alert to the Water Resources Authority (WRA) and local water management committees before the water is consumed.

The economic implications are profound. A single handheld EC meter costs a fraction of a full laboratory analysis, and a solar-powered telemetry buoy can operate autonomously for years. By replacing thousands of dollars of annual laboratory sampling with a network of automated EC probes, the WRA could increase its monitoring coverage by an order of magnitude while simultaneously reducing costs. Climate adaptation funds should therefore prioritize the procurement and deployment of these sensor networks, coupled with training programs to build local technical capacity for maintenance and data interpretation. The ML soft sensor is not merely a technological convenience—it is the prerequisite for timely, actionable intelligence in a region where water quality can change faster than traditional monitoring can respond.

Significance during insitu water quality tests in the Merti employing only the use of EC-meter or TDS-meter

The Merti Aquifer falls into three categories — proximal, peripheral, and distal — which respectively represent the lowest, medium, and highest levels of EC values. The machine learning outputs imply that if we know the EC in the portion of the Merti Aquifer being mapped, then we know the Sodium levels in that aquifer subset.

EC (electrical conductivity) is a bulk measure of how many dissolved ions are in the water. In the Merti Aquifer, the dissolved ions are dominated by **Na^+ and Cl^-** (this is what PC1 shows: Na, Cl, TDS, EC all load together). Because Na and Cl are the dominant charge pair, EC is effectively a **proxy for the Na–Cl concentration**. The ML model confirms this numerically: it can predict Sodium from EC (and TDS and Cl) with **$R^2 = 0.9856$** — meaning **98.6% of the variance in Sodium is already captured by EC, TDS, and Cl**.

Implication : a field technician with a $200 handheld EC meter can measure EC in 30 seconds and know, with ~98% confidence, whether the aquifer at that borehole has **high or low Sodium** — i.e., whether the flowing groundwater is fresh or saline. No laboratory, no flame photometer, no waiting weeks for results.

The Importance of Sodium Prediction in a Changing Climate

(a) Food Security and Small-Scale Farming as a Climate Resilience Strategy

As climate change continues to intensify droughts and erratic rainfall across the Horn of Africa, traditional pastoralism is becoming increasingly precarious. In response, both governments and development agencies are actively promoting a shift toward irrigated, small-scale farming (agro-pastoralism) as a primary climate adaptation strategy to secure food and income for rural communities. In the Merti region, this means utilizing the shallow groundwater of the aquifer to grow drought-resistant crops. However, the success of this adaptation strategy hinges entirely on water quality. If the water used for irrigation is highly saline, the transition to farming will fail. Therefore, accurately predicting Sodium levels using machine learning is not merely an academic exercise; it is a critical tool for national food security. It allows planners to identify which sections of the aquifer are safe for irrigation, ensuring that climate resilience investments are directed toward viable agricultural zones rather than being wasted on lands where the water will destroy the crops.

(b) The Permanent Destruction of Soil Structure by High Sodium

The stakes of accurate Sodium prediction are extraordinarily high because high-sodium water does not just harm the current crop—it permanently destroys the soil. When irrigation water contains excessive Sodium, a process known as "sodicity" occurs. The Sodium ions displace Calcium and Magnesium in the clay particles, causing the soil aggregates to disperse and collapse. This destroys the soil's structure, turning it into a dense, impermeable hardpan. Once this occurs, water and air cannot penetrate the soil, plant roots cannot grow, and the land becomes completely barren. This degradation is irreversible and has already rendered vast tracts of historically fertile land in arid regions entirely unsuitable for farming. Because the Merti Aquifer is highly vulnerable to evaporative concentration and saline up-coning, the risk of irrigating with high-sodium water is immense.

(c) Machine Learning as a Proactive Defense for Soil and Livelihoods

Given the irreversibility of soil sodicity, reactive monitoring is far too dangerous. By the time a farmer notices that their crops are failing due to saline water, the soil damage may already be permanent. This is where the ML-based diagnostic dashboard developed in this study becomes indispensable. By utilizing cheap, field-based Electrical Conductivity (EC) meters and the machine learning models (like the Stacking Ensemble and BPNN, which achieved near-perfect accuracy), water managers can instantly predict the Sodium concentration of a proposed well before a single drop is used for irrigation. This proactive approach allows agricultural extension officers to map safe irrigation zones, preventing the deployment of saline water onto fragile soils. Predicting Sodium levels is therefore the frontline defense against the permanent loss of arable land, safeguarding both the environment and the livelihoods of small-scale farmers navigating a changing climate.

2. Managed Aquifer Recharge (MAR) to Restore Flushing

The hydrochemical evidence presented in this study points to a fundamental diagnosis: the Merti Aquifer is salinizing because the flushing mechanism that historically pushed salts downstream has been severely weakened. The Ewaso Ng'iro River, which once flowed perennially to Sericho and into the Lorian Swamps, now dries up at Merti in most years [5]. The recharge that does occur is episodic, dependent on extreme flood events in the distant Kenyan highlands. This has created a dual problem: reduced recharge means less fresh water entering the aquifer, and reduced flow means less flushing of the saline water that already resides in the central corridor.

The solution is not to reduce abstraction alone—though that is necessary—but to actively restore the flushing mechanism through Managed Aquifer Recharge (MAR). Climate adaptation funding should prioritize the construction of sand dams, sub-surface dams, and recharge wells in the Isiolo-Meru transition zone, where the Ewaso Ng'iro crosses permeable alluvial deposits before entering the arid plains [1]. These structures would capture floodwaters that currently flow to the Indian Ocean unused, allowing them to infiltrate slowly into the aquifer rather than dissipating as surface runoff.

The hydrogeological rationale is supported by the PCA results. PC1, the salinity axis, is driven by evaporative concentration. By artificially augmenting recharge, the freshwater head in the aquifer would rise, creating a hydraulic gradient that pushes the saline interface back toward Somalia. This is the principle of hydraulic flushing: fresh water displaces saline water, just as a river pushes saltwater back at an estuary during high tide. The IAEA isotope studies confirmed that rapid recharge during historical flood events was the primary mechanism maintaining fresh water in the Merti beds [5]. MAR would artificially replicate this natural process.

Climate funds should also invest in water buffering technologies in the recharge zone, including infiltration basins and check dams along ephemeral wadis. The IGAD feasibility study identified these as priority interventions [1], but they have never been funded at scale. By implementing MAR upstream, the downstream communities in Garissa and Wajir would benefit from reduced salinity without having to reduce their own pumping—a politically and economically difficult intervention. MAR is therefore not just a technical fix; it is a transboundary equity mechanism that benefits the entire aquifer system. The Merti Aquifer's salinization is reversible if the flushing is restored, and MAR is the most direct way to achieve it.

3. Transboundary Governance Frameworks

The Merti Aquifer is not confined within Kenya's borders. Its flow terminates in Somalia's Jubbada Hoose region, meaning that every drop abstracted in Kenya reduces the flow available to Somali communities downstream. This transboundary reality creates a classic water-sharing dilemma: upstream users (in Kenya) have the physical advantage of extraction, while downstream users (in Somalia) bear the consequences of reduced flow and increased salinity. Without a governance framework, the tragedy of the commons is almost inevitable.

The hydrochemical and ML findings from this study provide the scientific evidence base needed to negotiate a transboundary agreement. The PCA biplot demonstrates that the aquifer's fresh water is a finite resource, and the ML model quantifies precisely how salinity responds to abstraction and recharge changes. This allows negotiators to move beyond vague assertions of "shared water" to concrete, data-driven scenarios. For example, the model can simulate the downstream salinity impact of a proposed wellfield in Wajir, or the benefit of an upstream MAR project in Isiolo. Such quantitative evidence is essential for building trust between riparian states.

The institutional vehicle for this governance already exists in embryonic form: the Dhobley–Liboi pilot project, facilitated by IGAD, which has been approved by Kenya and is awaiting approval from Somalia [1], [3]. This project represents a critical window of opportunity. Climate adaptation funds should be directed toward strengthening the institutional capacity of the IGAD Water Unit to serve as a neutral broker and technical secretariat. This would include funding for joint hydrogeological surveys, shared data platforms, and regular transboundary stakeholder forums. The ML dashboard developed in this study could be hosted by IGAD as a shared information resource, accessible to both Kenyan and Somali water managers.

Furthermore, the governance framework must address abstraction limits and wellfield spacing standards. The hydrochemical evidence of up-coning at Dhobley and Hoosingow—where continuous pumping for 24 hours has pushed EC above 3,000 µS/cm—demonstrates that current abstraction practices are unsustainable [1]. A transboundary agreement should include binding operational rules: minimum spacing between boreholes (e.g., ()1.5 km), maximum pumping rates, mandatory real-time EC monitoring, and seasonal abstraction limits during drought periods. These rules are not merely technical; they are political commitments that require sustained diplomatic engagement. Climate funds can provide the financial incentive for compliance, linking infrastructure financing to adherence to the agreed operational standards.

4. Operational Safeguards Against Up-Coning

Up-coning is the vertical migration of saline water from deeper formations into the freshwater lens, driven by the pressure differential created when a well is pumped. It is a well-documented phenomenon in coastal and arid aquifers worldwide, and the Merti Aquifer is no exception. The IGAD feasibility study documented that several existing boreholes—particularly at Dhobley and Hoosingow—are pumped at high yields continuously, penetrating the fresh-saline transition zone [1]. This creates a "cone of depression" around each well that draws saline water upward, permanently degrading the water quality at that location.

The operational safeguards against up-coning must be embedded in every aspect of wellfield design and operation. First, well placement should respect the geometry of the freshwater lens. Boreholes should be sited at a safe distance from the fresh-saline interface—typically at least 1.5 km—and the depth of the well screen should be carefully chosen to avoid penetrating the transition zone [7]. The hydrochemical data from this study provide a baseline for identifying these transition zones, and the ML models can predict Sodium concentrations at proposed well locations before drilling begins.

Second, pumping rates must be carefully controlled. The continuous 24-hour pumping observed at some existing wells is unsustainable. A sustainable pumping regime would involve shorter pumping cycles (e.g., 8–12 hours per day), with rest periods that allow the freshwater lens to recover. The ML model can be used to simulate different pumping scenarios and predict the resulting salinity changes over time, allowing water managers to set safe abstraction limits before degradation occurs.

Third, real-time monitoring is essential. Each wellfield should be equipped with in-line EC sensors that continuously measure the electrical conductivity of the pumped water. When the EC begins to rise, it is an early warning that up-coning is occurring. The ML model can then estimate the corresponding Sodium concentration and trigger an alert. This allows operators to reduce pumping immediately, before irreversible damage occurs.

Finally, wellfield management plans should include contingency measures for drought periods. During severe droughts, when recharge is minimal, the pumping rates should be reduced further to protect the lens. Climate adaptation funds should condition infrastructure financing on the adoption of these operational safeguards, ensuring that new wells are designed and operated sustainably from the outset. The Merti Aquifer's future depends not just on how much water is pumped, but on how intelligently it is pumped. The safeguards outlined here are the difference between a resource that lasts for generations and one that is lost within decades.

SUMMARY

The integration of Fluoride into this multivariate analysis is not merely a routine inclusion; it serves as the geogenic counterweight to climate-driven Sodium. By mathematically proving their orthogonality (Pearson r = 0.252), this study demonstrates that the Merti Aquifer faces a dual threat. Climate change is causing evaporative salinization (Sodium), while the natural geology of the recharge zone is leaching fluoride. This distinction is vital for policymakers: interventions to stop salinization (like MAR) will not solve the fluoride problem, and vice versa. Consequently, climate adaptation funding must address both the climatic and geogenic vulnerabilities of the aquifer."

This entire study is built on the argument that the aquifer is being degraded by two completely different forces: climate change (which drives Sodium up) and geology (which drives Fluoride up). If the study had only looked at Sodium, another study could argue that all the bad water is just coming from one single source (e.g., deep saline up-coning). By proving that Sodium and

Fluoride are orthogonal (independent), this study has proven that the aquifer has two distinct vulnerabilities that require two different management strategies, a Climate change variable(sodium) and a geologic variable (Fluoride) .

Fluoride acts as a geogenic fingerprint. Sodium and Chloride are "conservative" ions—they don't react much; they just concentrate when water evaporates. Fluoride, however, is a "geogenic" ion. It comes specifically from the dissolution of volcanic rocks (like the ones in the Mt. Kenya highlands). By tracking Fluoride, the study is actually tracking the geological "fingerprint" of the highland recharge. When Fluoride is low and Sodium is high, it is inferred that the water flow has stagnated in the lowlands and evaporated. This gives the researcher a powerful diagnostic tool to separate climate impacts from geological impacts.

REFERENCES

[1] IGAD Water Unit. (2024). Merti Transboundary Aquifer: Feasibility Study — Final Report, Volume I. IGAD Groundwater Information System.
[2] Water Resources Authority (Kenya). (2025). Exploratory Boreholes Hydrogeological Reports: Habaswein Site. Horn of Africa – Groundwater for Resilience Project (HoAGW4RP).
[3] IGAD. (2024). Merti Aquifer Final Report: Transboundary Development Proposals.
[4] Amimo, M. O., & Shune, J. A. (2022). Groundwater Sodium Levels Estimation of Proposed Irrigation Groundwater Source for the Kumahumato Settlement of the Dadaab Subcounty, North Eastern Kenya. International Journal of Progressive Sciences and Technologies, 32(2), 366–378.
[5] IAEA. (n.d.). Integrated Isotope and Conventional Hydrogeological Techniques for Managing Groundwater Resources in the Merti Aquifer. IAEA Technical Cooperation Project, Kenya.
[6] Water Resources Authority (Kenya). (2025). Hydrogeological and Geophysical Investigations: Hadado Site. HoAGW4RP.
[7] World Agroforestry Centre (ICRAF). (2019). Hydrogeological Risk Assessment of the Merti Aquifer: Impacts of Abstraction on Drawdown and Salinity. Acacia Water.
[8] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
[9] Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55–67.
[10] Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B*, 58(1), 267–288.
[11] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.
[12] Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. *ICLR 2015*.
[13] Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). Springer.
[14] Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259.
[15] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth.
[16] Fix, E., & Hodges, J. L. (1951). Discriminatory analysis, nonparametric discrimination: Consistency properties. *USAF School of Aviation Medicine*, Report No. 4.
[17] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.