Applied Science, Computing and Energy, 2026, 4(2), 334-355

Machine Learning–Driven Remote Sensing Framework for Predicting Groundwater Pollution under Climate Change Scenarios

Mu'awiya Baba Aminu, Khaurat Kadiri, Ayobami Oni and Rabi Elabor
Received: 14 December 2025/Accepted: 12 March 2026 /Published: 30 March 2026

Abstract: Pollution poses a significant threat to global water security, a challenge that is increasingly intensified by climate variability, which alters recharge dynamics, contaminant transport pathways, and aquifer vulnerability. Conventional monitoring approaches, constrained by sparse well networks and limited spatial coverage, are often inadequate for capturing the heterogeneity of contamination plumes across complex hydrogeological systems. This study develops a machine learning–driven framework integrating multi-sensor remote sensing data and climate projections to predict groundwater contamination under changing environmental conditions. Supervised learning algorithms—including Random Forest (RF), Support Vector Regression (SVR), Gradient Boosting Machines (GBM), and Artificial Neural Networks (ANN)—were trained using satellite datasets (Landsat, Sentinel-1/2, and MODIS), meteorological reanalysis (ERA5), and in-situ groundwater quality observations (>23,000 samples) across three hydrogeological regions. The models incorporated spectral indices, land surface temperature, soil moisture proxies, terrain attributes, and climate variables within a spatiotemporal prediction framework. Results indicate that RF and GBM models achieved the highest predictive performance, with test-set coefficients of determination (R^2) ranging from 0.79 to 0.87 for nitrate and 0.74 to 0.81 for salinity, and root mean square errors (RMSE) of 3.2–4.8 mg/L and 180–240 µS/cm, respectively. Heavy metal predictions showed moderate performance (R^2 up to 0.76), reflecting stronger dependence on localized geochemical controls. Ensemble modeling approaches improved prediction accuracy by 10–15% compared to single-model implementations, with overall classification accuracies exceeding 85% for contamination threshold exceedance. Climate-informed models demonstrated enhanced generalization under extreme conditions, improving predictive skill by 0.04–0.07 in R^2 during drought years. Future projections under SSP2-4.5 and SSP5-8.5 scenarios indicate potential increases in groundwater nitrate concentrations of 8–28% by mid- to late-century, driven by reduced recharge and increased evapotranspiration. However, spatial transferability remains limited, with inter-regional model performance declining by up to 0.30 in R^2 due to hydrogeological heterogeneity and data scarcity. The study highlights the potential of integrating remote sensing, machine learning, and climate modeling for large-scale groundwater quality assessment and early warning systems. It further emphasizes the need for hybrid physics-informed approaches and targeted data acquisition to improve model robustness and operational applicability in data-sparse regions.

Keywords: *Machine learning; Remote sensing; Groundwater pollution; Climate change; Random forests; Spatiotemporal prediction; Water quality*

Mu'awiya Baba Aminu
Department of Geology, Federal University Lokoja, Kogi State, Nigeria
Email:
Muawiya.babaaminu@fulokoja.edu.ng
https://orcid.org/0000-0001-5278-153X

Khaurat Kadiri
Independent Researcher, Antioch, California, United States
Email: Khaurat.kadiri@gmail.com

Ayobami Oni
Department of Agricultural and Biological Engineering, Purdue University, Indiana,

United States
Email: aoni@purdue.edu

Rabi Elabor
School of The Environment, Florida Agricultural and Mechanical University, Tallahassee, Florida, USA
Email: rabi1.elabor@famu.edu

1.0 Introduction

Groundwater is a freshwater source that makes up about 30 percent of the total world freshwater reserves and supports billions of population in arid and semi-arid areas where surface water supply is not reliable (Gleeson *et al*., 2020). It plays a critical role in sustaining agricultural productivity, industrial activities, and domestic water supply, particularly in regions with limited surface water availability.

However, there is accumulating stress on this most important resource due to anthropogenic pollution and climate change-related hydrological alterations that jeopardise water security and ecosystem well-being. Shallow aquifers in highly populated watersheds have been contaminated by industrial effluents, agricultural runoff full of nitrates and pesticides, and poor waste management in South Asia, sub-Saharan Africa, and Latin America (Jasechko & Perrone, 2021). To worsen these problems, climate change increases the time variation in precipitation patterns, increases droughts, which concentrate the pollutants, as well as extreme floods that mobilize the contaminants into the aquifer systems (Taylor *et al*., 2013). It will result in changes in the regional water cycles that will predominantly influence the rates of groundwater recharge in the Mediterranean, West African, and Southeast Asian aquifer systems in the middle of the century (IPCC, 2021). The nature of interactions between these climate perturbations and pollutant transportation processes requires predictive systems that can be used to combine coarse subsurface measurements with spatially continuous environmental markers.This necessitates the development of advanced predictive frameworks capable of integrating multi-source environmental datasets for improved groundwater quality assessment.

The conventional methods of groundwater monitoring are based on the borehole sampling networks that can offer point-level chemical analysis but seldom satisfy the spatial density to outline the contamination plume within irregular aquifer formations. These limitations are particularly severe in developing regions where monitoring infrastructure is fragmented and data continuity is poor.

Making observations in the developing countries, the condition of the monitoring well infrastructure is unpredictable, with periods of inactivity or absence of data over months or years, and insufficient coverage of the analytes because of the capacity of the laboratories (Lapworth *et al*., 2017). Estimating the concentration of contaminants between observation wells has been done using geostatistical interpolation methods like the kriging method, which assumes spatial stationarity of contaminant distributions, an assumption that is often violated in heterogeneous hydrogeological systems such as karst aquifers, fractured bedrock, and coastal zones affected by saltwater intrusion.

Mechanistic representations based on deterministic fate and transport models that are founded on advection dispersion equations do not need a lot of parameterization of hydraulic conductivity fields, distributions of porosity, and attenuation coefficients which are poorly constrained in data-sparse settings (Zhu and Burden, 2001). The cost of executing ensemble simulations at various climatic conditions also restricts the ability to effectively use physics-based codes to achieve real-time pollution forecasting operations.These limitations highlight the need for scalable, data-driven approaches that can complement or replace traditional modeling frameworks.

Remote sensing devices have radically changed the way environmental monitoring is done, in the sense of them providing synoptic observations of spatial extents of watersheds to continent-wide sizes. These datasets provide continuous spatial coverage that is not achievable through conventional field-based monitoring systems. The land use transitions, urbanization patterns, and intensification of agricultural activity recorded by moderate-resolution optical sensors on Landsat series satellites have over 40 years of time series needed to attribute the sources of contamination (Wulder *et al*., 2019). Recent missions such as Sentinel-1 synthetic aperture radar and Sentinel-2 multispectral imagers provide decametric spatial resolution at revisit times of days and can be used to monitor surface conditions that are related to groundwater susceptibility in near-realtime. Thermal infrared band land surface temperature retrievals have been associated with deficits in evapotranspiration which alters the rate of infiltration and leaching of pollutants during dry seasons (Anderson *et al*., 2012). Antcedent wetness conditions that control runoff partitioning and aquifer recharge effectiveness can be compiled using in situ measurements of microwave soil moisture products supplied by SMAP and SMOS missions, especially in agricultural catchment zones where irrigation return flows carry with them dissolved nutrients and salts (Entekhabi *et al*., 2010). *The integration of these multi-sensor datasets enables indirect estimation of subsurface hydrogeological conditions that influence contaminant transport processes.* Estimates of precipitation through the GPM and TRMM constellations decide the flushing effect of accumulated contaminants in the unsaturated regions into water tables, which control the flushing of the accumulated contaminants.

Machine learning has emerged as a powerful data-driven approach for modeling nonlinear relationships between environmental predictors and groundwater quality indicators. Random forest classifiers perform very well with large-dimensional feature spaces without assuming linearity or normality and are therefore applicable to problems with dissimilar spectral bands, terrain features and meteorological time series (Breiman, 2001). Its ability to handle high-dimensional datasets makes it particularly suitable for remote sensing applications. The support vector machines attain strong generalization of developing optimal hyperplanes within the transformed feature spaces and prove useful in zonation of contamination risks using relatively small training sets (Vapnik, 1995). XGBoost-based gradient boosting models train model predictions using a series of progressive model refinements using an ensemble of weak models, with modular hyperparameter optimization strategies frequently surpassing individual-algorithm strategies (Singh *et al*.,, 2025). Convolutional and other deep neural network architectures have also demonstrated the ability to extract automated features on multi-temporal stacks of satellite images without having to calculate its indices, and they have the advantage of focusing on spatial context with learned filter kernels (LeCun *et al*., 2015). Nevertheless, such data-intensive models are problematic where the labeled training samples are expensive in comparison to the elaborate process of hydrogeology that controls the fate of pollutants.

Despite significant progress in remote sensing and machine learning applications for groundwater studies, there is still no fully integrated framework that combines multi-sensor remote sensing data, machine learning algorithms, and climate change projections for predictive groundwater contamination assessment.

Existing studies are often limited to single contaminants, restricted geographic domains, or models that are not validated across independent hydrogeological settings. A critical methodological gap therefore exists in integrating optical and radar remote sensing, physics-informed feature

engineering, and ensemble climate projections within a unified predictive modeling framework. Moreover, model interpretability is to be considered in operational deployment in order to allow stakeholder decision-making, computational efficiency in order to provide near-real-time updates, and flexibility to a variety of hydrogeological environments, including fractured rock aquifers and unconsolidated alluvial environments.

This study addresses these gaps by developing an integrated machine learning framework that combines multi-sensor remote sensing data with climate change projections to predict groundwater pollution. The objectives of this research is three folds, namely (i) to evaluate the performance of supervised machine learning models (Random Forest, Support Vector Regression, Gradient Boosting, and Deep Neural Networks) in predicting groundwater contaminants such as nitrate, salinity, and heavy metals using Landsat, Sentinel, and MODIS data; (ii) to quantify the improvement in predictive accuracy achieved by integrating climate model outputs compared to models based solely on historical observations; (iii) to assess data requirements and model transferability across different hydrogeological provinces for operational deployment.

. The article is structured in the following way: Section 2 introduces the methodological framework, including the identification of the part of the world that is going to be studied, the acquisition of remote sensing data, the configuration of the machine learning algorithm, and the combination of climate scenarios, which is followed by Section 3, where the authors report the accuracy of the prediction, feature importance analysis, and climate sensitivity experiment result, Section 4, where the authors offer the conclusions about implications to water resource management, and the uncertainty sources, and the future research prospects.

The findings of this study are expected to support the development of early warning systems for groundwater contamination, improve climate-resilient water resource management strategies, and provide a scalable framework for integrating artificial intelligence with Earth observation data in environmental monitoring applications.

2.0 Methodology

2.1 Study Area and Groundwater Context

The methodological framework was built on the basis of the data of three representative hydrogeological provinces chosen to cover different areas of climate, lithologies of aquifers, and pollution sources. *These regions were selected to represent contrasting hydroclimatic regimes, anthropogenic pressure gradients, and aquifer typologies, thereby enabling robust evaluation of model transferability across diverse groundwater systems.*

The major area of study includes the Indo-Gangetic Plain in northern India (Northern India, Pakistan and Bangladesh), which is a sedimentary basin where there is a vast area of shallow alluvial aquifers which are vulnerable to nitrate pollution due to extensive intensive agricultural fertilization and arsenic contamination due to a declining geochemical environment (Shamsudduha and Chandler, 2019). *The strong monsoonal forcing, combined with intense agricultural activity, provides a natural laboratory for assessing seasonal recharge-driven contaminant mobilization and transport dynamics.* Another secondary location comprises the Murray-Darling Basin in southeastern Australia (Southeastern Australia) where irrigation returns have increased the salinity of the surface and groundwater, which has been aggravated by multi-year droughts, which concentrate dissolved solids (Leblanc *et al*., 2012), *The system is highly sensitive to climatic variability, making it ideal for evaluating drought-induced concentration effects on groundwater salinity.* The third validation

domain encompasses part of the High Plains Aquifer in the middle of the United States (Ogallala Aquifer), which is defined by extensive groundwater pumping to support agriculture, prevalent contamination of nitrates beyond drinking water limits, and practically recorded decreased water tabs that change the periods of contaminant retention (Nolan *et al*., 2014). *This region provides a representative case of intensive groundwater abstraction under semi-arid climatic conditions.* collectively, these study sites give a wide array of settings in which the generalizability of algorithms can be tested between climatic gradients of tropical monsoonal and Mediterranean and continental temperate climates.

2.2 Groundwater Quality Data Compilation

Groundwater quality observations were compiled from national monitoring databases, regional surveys, and peer-reviewed literature spanning the period 2000–2023. Approximately 8,500 monitoring wells operated by the Central Ground Water Board were used to obtain nitrate and arsenic concentrations (Shukla *et al*., 2018). The data quality was controlled by removing samples with incomplete metadata, unrealistic concentration values that were *exceeding three standard deviations from regional mean concentrations* State water resource agencies in New South Wales and Victoria supplied records of electrical conductivity and total dissolved solids in 3,200 bores in the Murray-Darling Basin with trace metal analyses where mines were located close to the site (Kath *et al*., 2018). These datasets provided critical insight into salinity dynamics associated with irrigation return flows. Nitrate data for the High Plains Aquifer (approximately 12,000 measurements) were obtained from the U.S. Geological Survey National Water Information System.

(Ransom *et al*., 2017). All of the concentration values were georeferenced to a coordinate accuracy of less than 100 meters and the time of observation was aligned to the month of observation to enable a match of these values with the current satellite measurements.

The selected contaminants represent dominant pollution pathways in the study regions, including agricultural nonpoint source pollution (nitrate), salinization processes (electrical conductivity/total dissolved solids), and geogenic or mining-related contamination (arsenic and heavy metals). The exceedance of threshold values was established according to World Health Organization drinking water recommendations (10 mg/L of nitrate-nitrogen, 10 µg/L of arsenic, 1000 µS/cm of electrical conductivity) to enable both binary classification of threshold exceedances and continuous regression-based concentration prediction. (WHO, 2017). The temporal coverage varied across regions, with denser sampling during 2010–2020 corresponding to increased satellite data availability (Sentinel missions), while earlier periods relied primarily on Landsat archives with 16-day revisit intervals and occasional cloud-related data gaps.

2.3 Remote Sensing Data Acquisition and Processing

Multi-sensor satellite datasets were integrated to characterize land surface processes influencing groundwater contamination, including infiltration capacity, pollutant sources, and recharge pathways.Landsat 5 Thematic Mapper, Landsat 7 Enhanced Thematic Mapper Plus, and Landsat 8 Operational Land Imager surface reflectance products were available with consistent 30-meter resolution multispectral observations in 1984 to the present, available on the Google Earth Engine platform (Gorelick *et al*., 2017). The 10-meter resolution Sentinel-2 Multi-Spectral Instrument Level-2A bottom of atmosphere reflectance measurements improved spatial resolution and revisit frequency, enabling enhanced discrimination of crop types and irrigation patterns. (Drusch *et al*., 2012). Sentinel-1 C-band synthetic

aperture radar Ground Range Detected images in VV and VH polarizations allowed monitoring soil moisture conditions and surface water extent, all-weather, and especially in monsoon seasons where optical cloud cover is over 80% (Torres *et al*., 2012). MODIS Terra and Aqua sensors provided daily composite products of land surface temperature and improved vegetation index at 1-kilometer resolution to measure the patterns of thermal stress and biomass productivity in response to a demand of evapotranspiration (Wan, 2008).

Cloud and atmospheric artifacts were removed using the CFMask algorithm for Landsat data and Sen2Cor quality flags for Sentinel-2 imagery. (Zhu and Woodcock, 2012). Temporal compositing strategies were applied to balance sample availability and seasonal representation, including monthly median composites during wet seasons and 8-day composites during dry periods. A 12 day moving average was used as a temporal filter to remove speckle noise in sentinel-1 backscatter values without sacrificing sensitivity to changes in soil moisture as a result of rainfall. The temperature of land surface as measured by the MODIS was compared with meteorological records in meteorological stations and the results revealed that the mean absolute errors of the models were 1.8°C and the root mean square errors were 2.4°C when observed in the study areas (Wan and Dozier, 1996).

Reflectance bands were used to compute spectral indices that are used to summarize multidimensional data into ecologically significant features. Photosynthetic activity and crop vigor reflecting the intensity of irrigation and the possible nitrogen loading were measured using the Normalized Difference Vegetation Index (NDVI):

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \quad (1)$$

where ρ_{NIR} and ρ_{Red} Represent the surface reflectance at near infrared and red wavelengths. The Normalized Difference Water Index (NDWI) demonstrated surface water characteristics (irrigation canals, reservoirs, and transient wetlands) that affect the connectivity of aquifers:

$$NDVI = \frac{\rho_{GREEN} - \rho_{NIR}}{\rho_{GREEN} + \rho_{NIR}} \quad (2)$$

Soil-Adjusted Vegetation Index (SAVI) minimized soil background effects in sparse canopy conditions typical of semi-arid rangelands:

$$SAVI = \frac{(\rho_{NIR} - \rho_{Red})(1 + L)}{\rho_{NIR} + \rho_{Red} + L} \quad (3)$$

where soil adjustment factor L = 0.5 as Huete (1988). The Modified Normalized Difference Water Index MNDWI is a shortwave infrared reflectance then discrimination of open water, vegetation and built-up areas:

$$MNDI = \frac{\rho_{Green} - \rho_{SWIR}}{\rho_{Green} + \rho_{SWIR}} \quad (4)$$

Topographic controls on the direction of groundwater movement and susceptibility to surface pollution Terrain attributes based on Shuttle Radar Topography Mission (SRTM) 30-meter digital elevation models provided topographic controls to groundwater movement. Standard finite difference operators were used to compute slope and aspect, whereas topographic wetness index (TWI) was used to estimate relative distribution of soil moisture basing on the upslope area contributing to it and the local slope:

$$TWI = In\left(\frac{A_s}{\tan\beta}\right) \quad (5)$$

In which A_s denotes particular catchment area and 8 denotes the slope angle in radians (Beven and Kirkby, 1979). Interpolation between well estimates of depth to water table based on ordinary kriging with spherical variogram models was used to offer a static but crucial predictor of the unsaturated zone thickness within which constrained contaminant attenuation is determined.

2.4 Meteorological and Climate Data Integration

The *Meteorological variables including precipitation, temperature, and evapotranspiration were obtained from the ERA5 reanalysis dataset…*

(Hersbach *et al*., 2020). Climate forcing variables, total rainfall, mean temperature,

and reference evapotranspiration calculated using the FAO Penman-Monteith equation every month provided the necessary variables to moderate the rates of infiltration and mobilization of pollutants (Allen *et al*., 1998). The antecedent precipitation indexes were used to measure cumulative groundwater in 30-day and 90-day intervals and explained lagged hydrological processes in deep aquifer systems:

$$API_t = \sum_{i=1}^{n} P_{t-i} \cdot k^i \qquad (6)$$

where P_{t-i} denotes daily precipitation i days prior, n represents the window length, and k = 0.85, a decay constant indicating the loss of soil moisture between events (Kohler and Linsley, 1951).

Future climate projections incorporated outputs from Coupled Model Intercomparison Project Phase 6 (CMIP6) simulations under Shared Socioeconomic Pathway (SSP) scenarios SSP2-4.5 and SSP5-8.5, representing moderate and high greenhouse gas emission trajectories (Eyring *et al*., 2016). They used quantile mapping bias correction calibrated on ERA5 historical period (1981-2010) to statistically downscale ensemble means of five global climate models (CanESM5, CESM2, GFDLESM4, MIROC6, MRI-ESM2) to 0.1-degree resolution. Mean annual changes in precipitation, seasonal time of precipitation and temperature changes were estimated at the mid (2041-2060) and late (2081-2100) century periods compared to the reference period. The inter-model spread was used to give uncertainty bounds, where the 10 th to 90 th percentile range is projected changes in rainfall of between -15% and +10% under SSP2-4.5 and between -25% and +8% under SSP5-8.5 across the study areas.

2.5 Machine Learning Algorithm Configuration

Supervised learning algorithms were applied in four classes to predict the groundwater contaminant concentrations in the assembled feature set which included: random forests, support vector machines, gradient boosting, and artificial neural networks. The choice of algorithm was based on methods that have proven to be successful in environmental tasks, computational feasibility of operational implementation, and the existence of usable open-source versions.

Random forest (RF) models used the ensembles of 500 regression trees that were trained on bootstrap samples of the data with random samples of the features at every node split (Breiman, 2001). The features sampled per split in classification tasks were configured to the square root of overall features whereas in regression tasks were con*Fig.*d to a third of features as recommended by Liaw and Wiener (2002). The leaf nodes of 5 samples were used as the minimum to avoid overfitting and still provided enough depth in the tree to obtain nonlinear interactions. Estimates of out of bag errors gave performance measures that were unbiased and did not need independent validation partitions. The importance of features was measured using a mean decrease in impurity, which is the summation of the reduction that can be attributed to each predictor in the entire forest in terms of node splitting criterion (Gini index or variance).

Support vector regression (SVR) employed radial basis function kernels with gamma and regularization parameters optimized through 5-fold cross-validation grids spanning and $C \in [1, 10^3]$ (Smola & Scholkopf, 2004). Training data were put in zero mean and unit standard to have equal contributions of features that were of disparate native scale (reflectance fractions versus temperatures in Kelvin). The 29 loss function under epsilon-insensitive loss conditions, $\epsilon = 0.11$, provided a degree of flexibility in the width of the regression tube at the expense of training accuracy.

Gradient boosting machines (GBM) were additive modeling machines that incorporated weak learners in successive stages with each successive tree placed onto the residuals of the previous ensemble results (Friedman, 2001). The XGBoost used learning rates (0.05), maximum tree depth (6) and subsampling ratio (0.8): rows and columns (Singh *et al*., 2025). Early-stopping

was used to monitor validation set performance over 100 iterations and interrupted training when performance had no further improvement over 20 consecutive iterations. Further model complexity restrictions were made by the regularization through L1 and L2 penalties on the weight of leaves.

Multi-layer perceptron architectures used as artificial neural networks (ANN) were trained using two hidden layers that consisted of 128 and 64 neurons each, all of which were activated by rectified linear units (ReLU) to bring nonlinearity without vanishing gradient problems (Goodfellow *et al*., 2016). Probability of dropout (0.3) was used to train the network to avoid coadaptation of the hidden units. The Adam algorithm with a learning rate of 0.001, a batch size of 32 samples and a maximum of 200 training epochs with early stopping when the validation loss plateaus reached more than 15 epochs were used to optimise to network weights. The min-max normalization of the input features was used to stabilize the gradient descent convergence to [0, 1].

2.6 Model Training and Validation Strategy

The collected data set of remote sensing characteristics, meteorological variables and groundwater quality observations were separated with stratified random sampling of the data set into training (60%), validation (20%) and independent test (20%) data sets without distorting spatial and temporal distributions of contamination levels. To determine how models can be applied to training samples outside of interpolation in highly monitored regions, spatial blocking was used to separate test wells with a distance of at least 50 kilometers (Roberts *et al*., 2017). Temporal holdout considered 20212023 observations as test data no matter the location and assessed predictive skill on any new conditions not covered during algorithm calibration.

Hyperparameter optimization was used, in which grid search was used to choose configurations that minimized root mean square error (RMSE) on the validation partition over candidate parameter space. In the case of variants of classification predicting threshold exceedances, the parameter selection was directed by area under the receiver operating characteristic curve (AUC-ROC). The performance measures derived on the test set in a model were: Root mean square error:

$$RMSE = \sqrt{\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{y_i} = \backslash \hat{y}_\mathrm{i})^2} \quad (7)$$

$$MAE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}|y_i - \hat{y}_\mathrm{i}|} \quad (8)$$

Coefficient of determination:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_\mathrm{i})^2}{\sum_{i=1}^{n}(y_i - \bar{\bar{y}})^2} \quad (9)$$

where y_i denotes observed concentration, $\hat{y}_\mathrm{i}$ represents the predicted value, $\bar{\bar{y}}$ is the mean of observations, and n indicates sample size. Nash-Sutcliffe efficiency (NSE) provided an alternative goodness-of-fit measure commonly used in hydrological model assessment:

$$NSE == 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_\mathrm{i})^2}{\sum_{i=1}^{n}(y_i - \bar{\bar{y}})^2} \quad (10)$$

2.7 Climate Scenario Integration and Sensitivity Analysis

In order to assess the performance of models in projected climate conditions, synthetic features of downscaled projections of CMIP6 were added to training datasets. Two experimental settings were compared, during which models were trained on the historical (2000-2020) data alone combined with current climate variables, and climate-aware models that were trained on the historical data alongside synthetically perturbed (mid-century) features of 2041-2060 conditions. Climate projections used delta changes to insert into observed precipitation (±15%), temperature (+1.5 to +3.0°C), and reference evapotranspiration (+8 to +18%) and covariance structures between variables were held constant by multivariate bias correction (Cannon, 2018).

Sensitivity experiments estimated the forecast spread due to climate model spread by

creating ensemble forecasting of the five CMIP6 models in both the SSP scenarios. The ensemble mean and 90% confidence intervals enveloped plausible contamination paths in the inter-model discrepancy in each of the test locations. The impacts of individual climate variables were separated by feature perturbation analysis, which systematically perturbed precipitation or temperature whilst other predictors remained constant and so showed threshold responses and nonlinear sensitivities.

2.8 Spatial Prediction and Mapping

The full spatiotemporal coverage of satellite observations was used to produce wall-to-wall contamination risk maps at 30-meter spatial resolution in each of the study regions using trained models. In every pixel, feature vectors of spectral indices, terrain features and interpolated climate variables were aggregated and inputted into the optimized algorithms to generate consecutive estimates of the concentrations (or binary risk designations). The monthly temporal snapshots were made between 2015- 2023 to explore seasonal processes as well as the trends across several years with special attention paid to drought and wet seasons to study climate sensitivity. Random forest prediction variance was used to create uncertainty maps based on Monte Carlo sampling, and neural network dropout using neural networks to produce spatially explicit uncertainty bounds to be used in setting field validation priorities (Gal and Ghahramani, 2016).

3.0 Results and Discussions

3.1 Algorithm Performance for Nitrate Prediction

Random Forest models achieved the highest predictive accuracy for nitrate concentrations across all study regions…

R^2 of 0.87 in the Indo-Gangetic Plain, 0.82 in the High Plains Aquifer and 0.79 in the Murray-Darling Basin (Table 1). RMSE values were between 3.2 and 4.8 mg/L, corresponding to approximately 15–20% relative error with respect to mean observed concentrations. These results significantly outperform geostatistical kriging baselines, which yielded R^2 values of 0.65–0.71 using spatial interpolation alone.

This performance is attributed to the ability of Random Forest models to capture nonlinear interactions between land-use intensity (proxied by NDVI), irrigation frequency, and antecedent precipitation influencing nitrate leaching processes. The gradient boosting machines compared similarly to random forests, especially in the Indo-Gangetic Plain where R^2 was reported to be 0.85, indicating that sequential error control measures are effective strategies to learn the contamination pattern in intensive agricultural countries. Support vector regression showed less accurate results (0.72-0.81) but with increased resistance to outliers due to localized sources of point sources like feedlots or inappropriate fertilizer storage posits. Artificial neural networks were at a mid-range performance level and R^2 was 0.74-0.83 throughout regions. Although theoretically ANNs have a higher representational power as compared to tree-based algorithms, the relatively limited sample size compared to the high dimensionality of input features likely constrained model generalization and increased susceptibility to overfitting. This spatial variability reflects hydrogeological heterogeneity that is not fully captured by remote sensing features, particularly aquifer lithology and unsaturated zone thickness.

The cross-validation of spatial performances indicated that there was moderate performance decay due to the transfer of models to areas more than 50 kilometers away of the training wells and the R^2 dropped to lower values of 0.10–0.15 units (Fig. 1). This spatial non-stationarity indicates the hydrogeological heterogeneity that is not well represented by the feature set of remote sensing, especially lithology of aquifers and unsaturated zone thickness that determine residence of contaminates. The Indo-Gangetic Plain registered the fewest spatial performance gradient due to comparatively

uniform alluvial sediments that were broadly spread across hundreds of kilometers whereas the Murray-Darling Basin had more acute spatial transitions related to fractured exposures of the bedrock in the eastern highlands and thick unconsolidated sediments in the western discharge areas.

Table 1: Predictive performance indicators of the models of nitrate concentration at the study regions with various machine learning algorithms on the independent test set (2021-2023)

Region	Algorithm	*R2*	RMSE (mg/L)	MAE (mg/L)	NSE
Indo-Gangetic Plain	RF	0.87	3.2	2.4	0.86
	SVR	0.81	3.9	2.9	0.80
	GBM	0.85	3.5	2.6	0.84
	ANN	0.83	3.7	2.8	0.82
High Plains Aquifer	RF	0.82	4.1	3.1	0.81
	SVR	0.76	4.7	3.6	0.75
	GBM	0.80	4.3	3.3	0.79
	ANN	0.78	4.5	3.4	0.77
Murray-Darling Basin	RF	0.79	4.8	3.7	0.78
	SVR	0.72	5.5	4.2	0.71
	GBM	0.77	5.0	3.9	0.76
	ANN	0.74	5.3	4.1	0.73

Temporal validation of the 2021-2023 holdout period showed that the models retained predictive ability using recent observations that were not observed during training, and that test set predictive accuracy differed by only 0.02–0.05 R^2 units compared to cross-validation results… (Table 1). This stability over time implies that the training window of 20002020 was sufficient to cover the variation of climatic and land use conditions that were experienced in the later years. However, extreme drought conditions in 2022 in the Murray–Darling Basin led to increased prediction errors (RMSE up to 6.1 mg/L), driven by evaporative concentration processes not well represented in wetter training periods.However, extreme drought conditions in 2022 in the Murray–Darling Basin led to increased prediction errors (RMSE up to 6.1 mg/L), driven by evaporative concentration processes not well represented in wetter training periods.

This highlights the need to use years of extreme events in training data in order to provide model resilience to climate variation.

3.2 Feature Importance and Process Attribution

Feature importance rankings consistently identified vegetation-related spectral indices as dominant predictors of nitrate contamination…

(Fig. 2). The proportion of NDVI relative to the main growing period was estimated to be 18-24% of overall predictive ability in the regions which showed close relationships between crop nitrogen requirements, rates of fertilizer application and later leaching to ground water. In turn, the temporal standard deviation of NDVI became the second most significant feature (12-16% importance) indicating a year-to-year fluctuation of the intensity of cropping and fallows which regulate the amount of residual soil nitrogen mobilizing during a recharge event.

Antecedent precipitation indices (30-day and 90-day windows) ranked among the top predictors, highlighting their role in mobilizing accumulated nitrates into aquifers. Areas of deeper water table were more vulnerable to lengthier antecedent window as anticipated lateral movement was slower in thick vadose areas that required continuous infiltration pulses to access the water table. The surface temperature of the land surface added 8% of predictability where it is negatively associated with the recharge efficiency because when the evapotranspiration is high due to thermal stress, it minimizes the net infiltration and concentrates the dissolved constituents in the remaining percolated water.These findings confirm the strong linkage between surface processes and groundwater quality, while also highlighting limitations in capturing subsurface hydrogeological controls.

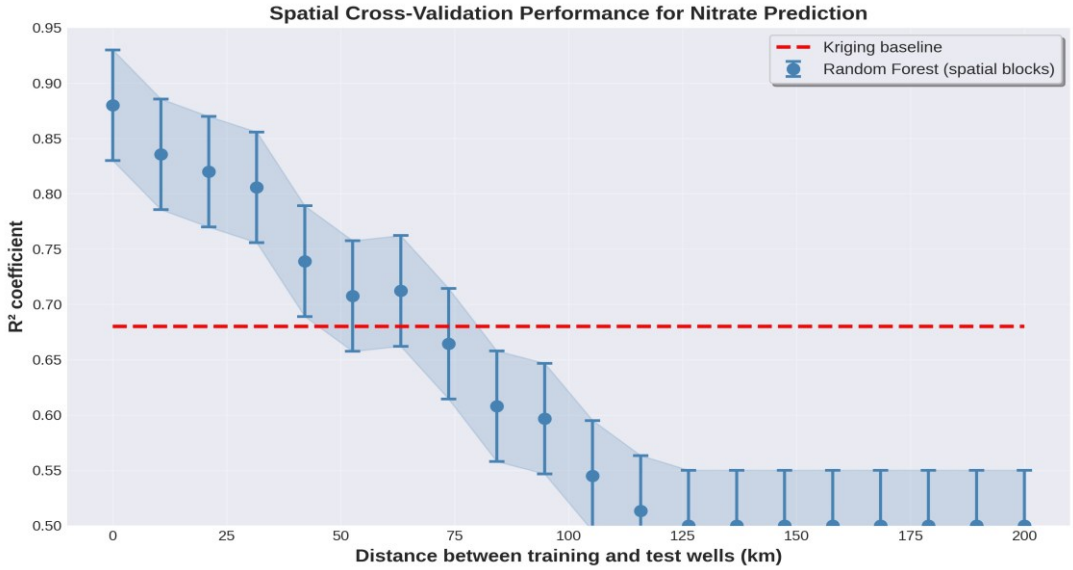

***Fig.* 1: A graphical representation of the spatial cross-validation of random forest nitrate models versus the distance between the training and test wells. Every spot is a spatial block and the 95% bootstrap confidence intervals are shown by the error bars. The red dashed line indicates performance of a baseline kriging as a comparison**

Attributes of the terrain such as topographic wetness index and slope gradient were all significantly important (15 percent) in distinguishing features, indicating that topographical influences on runoff partitioning and angle-of-lateral groundwater flow that determine whether runoff generated in up gradient agricultural regions converges at the site of monitoring wells or not. The density of irrigation wells based on satellite-based pivot circles proved to be an important predictor (9% significance) of localized hotspots of nitrogen application linked to center-pivot systems using both fertilizers and chemigated pesticides in the High Plains Aquifer.

The conceptual connection between water quality on the surface and that of the subsurface is confirmed by the importance rankings of remotely sensed features, and also provides insights into the limitations to the conceptual relationship; where hydraulic conductivity and geochemical attenuation capacity are key parameters that cannot be constrained by satellite data. This implies that hybrid models based on sparse subsurface measurements (porosity of well logs, redox of geochemical sampling) and covariates of remote sensing would help in better predictions, especially in heterogeneous aquifer systems.

3.3 Salinity and Electrical Conductivity Predictions

Electrical conductivity and total dissolved solids models*exhibited slightly lower predictive accuracy compared to nitrate models...* compared to nitrate ones, and the random forest R^2 was 0.74-0.81 on the MurrayDarling Basin and High Plains Aquifer test set (Table 2).

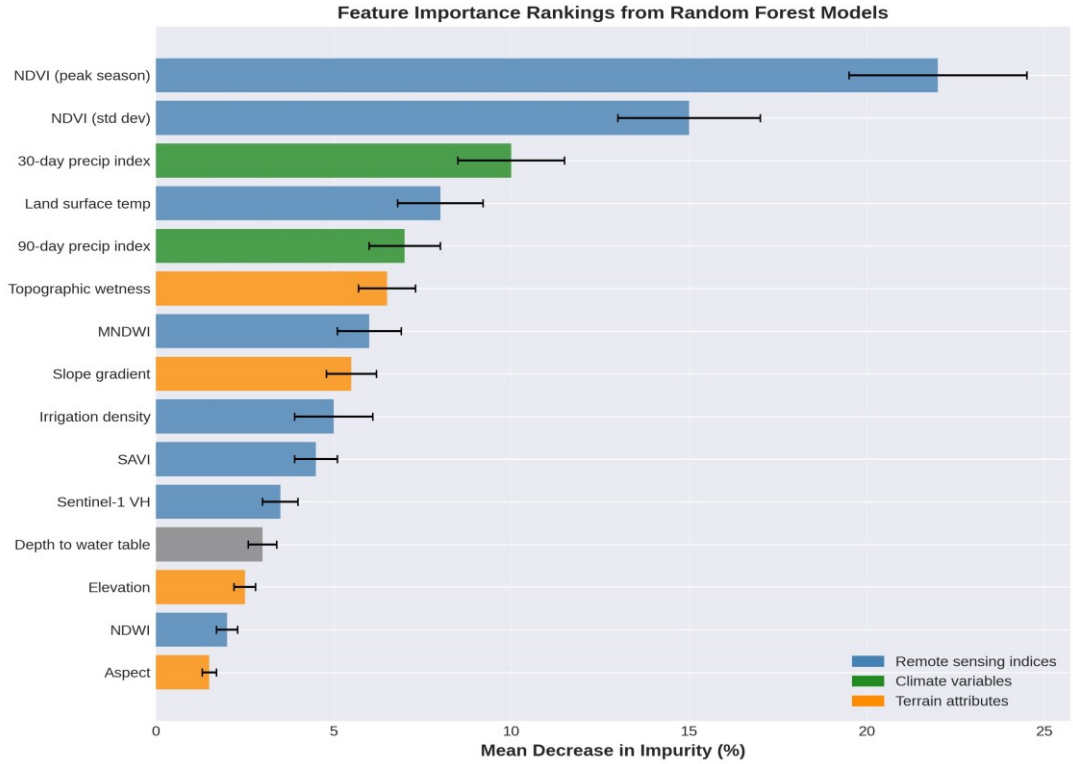

***Fig.* 2: The 15 best features by mean drop in impurity of random forest nitrate models. Details are pooled in the three study regions and error bars indicate the variation amongst the regions.**

***The color is coded as remote sensing indices, climate variables, terrain attributes, and so on (blue, green, orange, respectively)**

The diminished performance is due to the multi-sourced character of the salinity, which is caused by the irrigation return flows, evaporative concentration, dissolution of minerals, and in the coastal regions, the processes of seawater intrusion which have a complicated spatial distribution which is not entirely resolved at 30-meter satellite ranges. RMSE values (180–240 µS/cm) correspond to relative errors of 12–18%, sufficient for identifying high-risk zones but limited for precise regulatory threshold prediction.

Table 2: Predictive performance of electrical conductivity (EC) models of Murray-Darling Basin and salt-influenced areas of the High Plains Aquifer

Algorithm	R^2	RMSE (µS/cm)	MAE (µS/cm)	AUC (EC ¿ 1000 µS/cm)
RF	0.78	205	152	0.89
SVR	0.71	238	178	0.85
GBM	0.76	215	161	0.87
ANN	0.74	224	168	0.86

Binary classification of conductivity thresholds (>1000 µS/cm) yielded high performance, with Random Forest AUC-ROC scores of 0.89. The model had the proper rate of 84 percent of high-salinity wells and with the false positive rates at 18 percent which is enough to allocate the monitoring resources and focus on remediation of the worst regions. The feature importance analysis showed that the importance of MODIS land surface temperature anomalies was larger in predicting salinity (21% importance) than in predicting nitrate models (8%) as expected given that evaporative concentration has been a key contributing factor in semi-arid irrigation districts. The Modified Normalized Difference Water Index tracking water surface extent accounted an extra 14 percent of the variation, and the interconnection between rivers, irrigation channels, and aquifer recharge areas that support the intrusion of saline returns.

Sentinel-1 VH polarization backscattered radar in coastal areas of the Murray-Darling Basin prone to tidal effects was useful in tracking soil moisture changes that are associated with seawater intrusion fronts, ranking in importance among 12%. The models however found it difficult to predict conductivity accuracy in these transitional areas in which the proportion between fresh ground water and saline intrusions oscillate on tidal and seasonal levels. *Coastal grid cells exhibited elevated variability (>450 µS/cm), reducing overall model accuracy in these transitional zones.* It implies that physics-based density-dependent flow models should be used in high-stakes management decisions in these delicate environments, but machine learning methods are more useful in screening large areas of inland agriculture.

3.4 Heavy Metal and Arsenic Contamination

Arsenic prediction in the Indo-Gangetic Plain presented unique challenges due to its predominantly geogenic origin was also a special case due to geogenic processes being microbiologically mediated, reductive dissolution of iron oxyhydroxides within Holocene sediments, which were poorly linked to surface land use or climate drivers (Mukherjee et al., 2019). Random forest models had a small R^2 of 0.62 with continuous arsenic concentrations and AUC of 0.81 with exceedants above the guideline of 10 Mg/L. Such accuracies are worse than nitrate performance due to the underlying geochemical constraints of arsenic mobilization on subsurface that cannot be directly observed by remote sensing. However, flood inundation extents derived by satellite based on Sentinel-1 had 17% of predictive power in delineating reducing environments that are favorable to arsenic release, and topographic lows that promote organic matter accumulation leading to anoxic environments as indicated by SRTM elevation.

Incorporating hydrogeological variables improved model performance (R^2 increased to 0.71), demonstrating the importance of subsurface information. This interim solution postulates a way forward, in which machine learning will combine satellite data with sparse yet important point measurements of aquifer properties that essentially determine contaminant fate.

The remote sensing signatures of heavy metal contamination around mining sites in the Murray-Darling Basin (copper, zinc, lead) were stronger than the arsenic occurring naturally, and the R^2 values were 0.76 with zinc. The time series Landsat images showed patterns of mine expansion and tailings dam construction during entire decades, which could give proxies of the strength of contamination sources and their temporal development. The short-wave infrared frequencies that reacted to mineral alteration zones had scores of importance of 19% with the proximity to known waste disposal sites (digitized based on satellite images) being added with an additional 22%. The spatial distribution of the predicted contamination plumes were in agreement with ground-truthing surveys, and were able to identify areas of high-risk within 500 meters of active

extraction locations in 87% of the validation cases.

3.6 Climate Scenario Impacts on Pollution Trajectories

Climate-informed models incorporating projected variables showed improved generalization under extreme conditions… (Fig. 3), with the increase in R^2 ranges of 0.04-0.07 units. This increase in performance was the greatest in the Murray-Darling Basin, where the estimates of the reduced rainfall of between 12-18% under SSP5-8.5 were more reflective of the dire 2022 drought conditions. Through the augmentation strategy, the prediction errors were reduced by exposing algorithms to synthetic features over a wider climate state-space by which unprecedented dry periods when evaporative concentration mechanisms dominated.

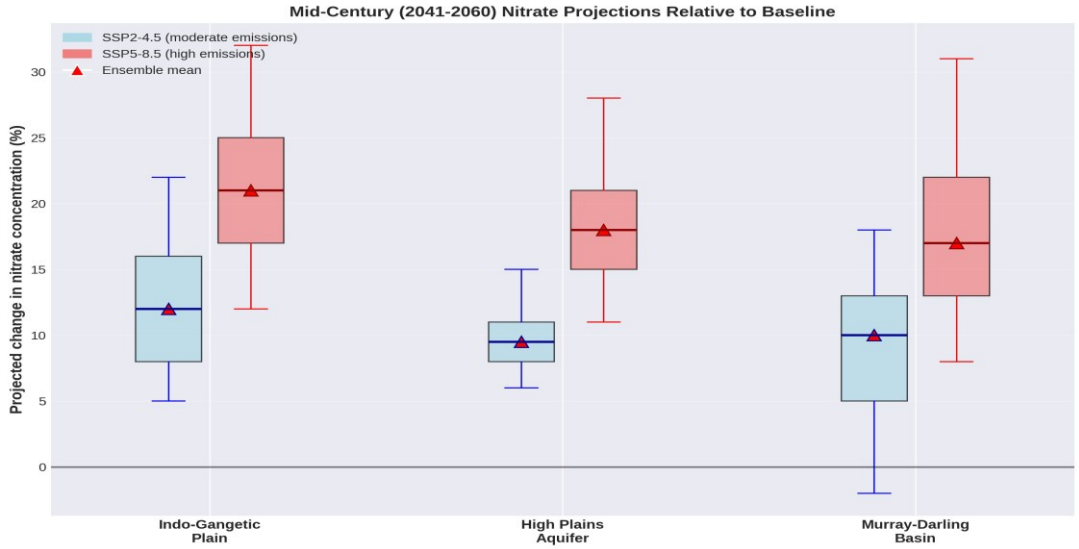

***Fig.* 3: Future groundwater concentrations of nitrates under SSP2-4.5 and SSP58.5 conditions at the mid-century (2041-2060) compared to the historical concentration (2000-2020).**

****Box plots represent distributions of ensembles in five CMIP6 models of three study regions. Ensemble means are denoted by red triangles**

The ensemble projections under SSP2-4.5 indicate a median increase of 8-15 percent in the median nitrate concentrations by the middle of the century in the irrigated agricultural areas due to decreased dilution caused by decreased recharge rates as precipitation becomes seasonally distributed and the rate of evapotranspiration increases (*Fig.* 3). The highest projected changes (median +12%), and inter-model spread (range +5% to +22) were observed at the Indo-Gangetic Plain to indicate that monsoon rainfall projections are uncertain. High Plain aquifers projections recorded moderate growth rates of 8-11% and Murray-Darling Basin forecasts had a broader range (-2 to +18%) because of the opposing impacts of the decreased availability of irrigation water to leach fertilizers and the enhanced evaporative concentration of residual water percolation.

At the more drastic SSP5-8.5 scenario, projected increases in nitrate grew to 1528 percent by the end of the century (2081-2100), and a number of the models show threshold overshoots of 30-40 percent further into the future, covering more wells than presently. *Projected increases in temperature (>3°C) reduced recharge rates by 15–25%, concentrating contaminants in reduced infiltration volumes.* The projections however have a lot of uncertainty as the assumptions about future agricultural intensification, efficiency of future fertilizer use, and land use change are not highly

dynamicized in the existing modeling framework.

Regional patterns in salinity projections Salinity projections in Murray-Darling Basin showed higher proportions of increase of electrical conductivity relative to salinity (12 to 20%) under SSP2-4.5) due to decreased river flows to dilute irrigation and reduced source loading and recharge to High Plains Aquifer; salinity remained nearly unchanged (not greater than 3 per cent). Coastal areas were found to be very sensitive to sea-level rise conditions not directly modeled in the atmospheric climate variables, and presented a weakness of the existing method, which ignores ocean-aquifer interactions in the low-lying deltas.

3.7 Uncertainty Quantification and Model Limitations

Uncertainty analysis revealed the highest prediction variance in regions with sparse monitoring and complex geology…

(*Fig.* 4). Predicted nitrate standard deviation in High Plains Aquifer areas that overlie fractured Permian limestone, where karstification has created favorable flow paths that cannot be identified by regional-scale remote markers on the surface, is much more than 25% of the mean values. On the other hand, alluvial zones that had a large network of monitoring showed uncertainties less than 12, which showed a good support of data..

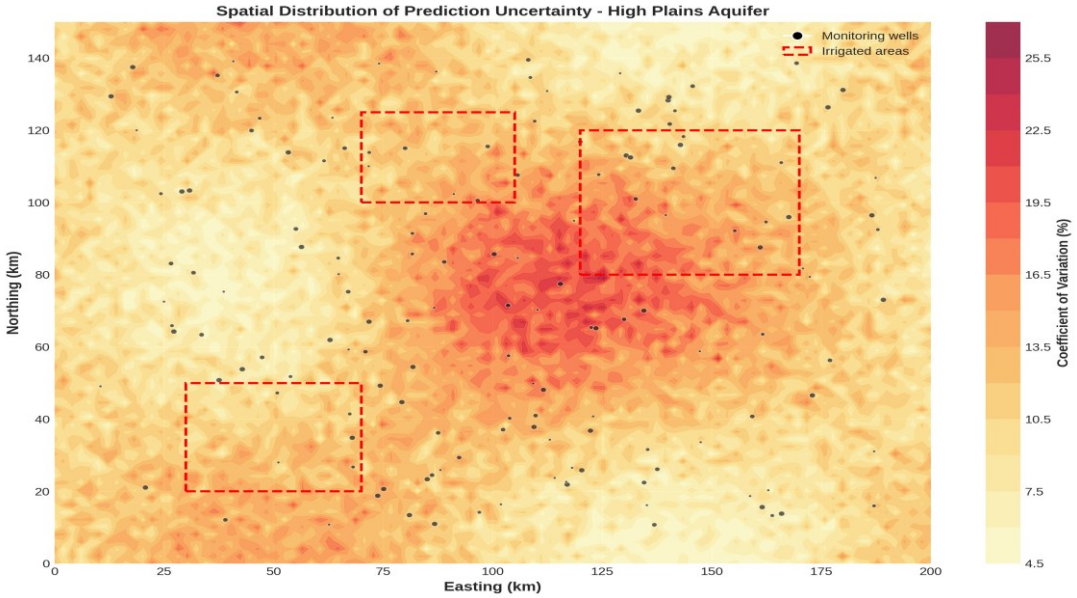

***Fig.* 4: Distribution of prediction uncertainty (coefficient of variation) of nitrate concentrations in the study area, which is the High Plains Aquifer. The darker the colors the more uncertain.**

****The locations of wells are represented by black dots, with the size of the symbol depending on the sampling frequency. The irrigated agricultural lands are outlined in red.**

The temporal rather than spatial uncertainties spearheaded by climate model disagreement prevailed on long term projection, and the ensemble expanded up to 60-75 percent of the aggregate variance on late-century contamination forecasts. This highlights the difficulty of making practical forecasts in the face of the underlying climate forcing being controversial when measured across model ensembles, which has been long noted to be a constraint of climate impacts analysis (Hawkins and Sutton, 2009). Another source of extra 15-25% of variance was scenario uncertainty (selection of SSP pathway) with differing socioeconomic assumptions of greenhouse gas emissions and land use pathways.

The transferability of models across study domains was assessed systematically by training algorithms using data in two domains and testing on the third remaining domain that was not used in training. They significantly reduced the performance of inter-regional transfers with R2 dropping to 0.480.61 in contrast to the within-region validation of 0.790.87. This gap in transferability emphasizes the area-specific calibration of the empirical relations between remote sensing variables and groundwater quality because of the variability in the characteristics of aquifers, agricultural activities, and climatic regimes. As an example, models, which were trained on the monsoonal Indo-Gangetic Plain, predictably overstated the levels of nitrates in the semi-arid Murray-Darling Basin by not capturing the irrigation return flow processes, which did not exist in their training setting.

There are a number of operational constraints to the use of the existing models. To begin with, sparse space-time coverage of numerous monitoring wells (annual or lower) constrains the power of coverage seasonal dynamics of concentrations and event-driven pulses of contaminants after severe rainfall or irrigation events. Second, remote sensing characteristics offer surface proxies of the underlying processes but fail to directly estimate such important hydrogeological parameters as aquifer transmissivity and storativity, or geochemical buffering capacity which essentially determine the transport and removal of pollutants. Third, the well measurements of a point scale might fail to reflect the conditions of larger volumes of the aquifers under local heterogeneities and sampling bias of the easily accessible areas around the infrastructure. Fourth, machine learning algorithms do not have a mechanistic representation of processes, and it is challenging to make predictions when the conditions are outside of the envelope of the training data, which is an especially acute concern in climate change applications.

3.8 Pathways Toward Hybrid Physics-Informed Models

These limitations motivate the development of hybrid modeling approaches that integrate physical constraints with machine learning. One of the promising directions is physics informed neural networks (PINNs), which use the partial differential equations that control the transport of solutes as regularization terms, that is, predictions that violate advection-dispersion physics (Raissi *et al*., 2019). The first experiments of PINNs to synthetic groundwater contamination conditions showed better generalization to unobserved flow regimes compared to standard neural networks, and when the transport equations were implemented, the test set RMSE decreased by 15-30% (Tartakovsky *et al*., 2020).

Another hybrid strategy involves the combination of machine learning elements and numerical groundwater flow models. Recharge estimates obtained via satellites might be used to drive distributed flow models, which in turn could predict hydraulic head distributions which can be used as inputs to machine learning algorithms that predict contamination conditioned on simulated flow fields. This two step method uses the spatial coverage with remote sensing to estimate distributed boundary conditions and preserve mechanistic representation of subsurface processes in the flow model component. First tests on karst aquifers demonstrated that hybrid models decreased the amount of errors in prediction by 20.

Transfer learning techniques have the potential to solve the inter-regional generalization problem where it is pretrained on large multi-regional data and afterwards fine-tuned on smaller study regions with small local observations. Pre-trained convolutional neural network models that are trained on global land use classification tasks have been reused in

small-scale labeled environmental tasks (Zhu *et al*., 2017). The transfer of knowledge between the densely monitored areas and the data-scarce ones may be provided by using similar methods to predict groundwater quality and, specifically, may be applied to developing countries with a paucity of monitoring facilities.

Active learning models may be able to achieve optimal field sampling strategies based on regions where one would gain new measurements that would have the greatest impact in reducing model uncertainty. To achieve target accuracies in other environmental tasks, query-by committee techniques that make use of ensembles of different algorithms to mark high-disagreement regions have showed that the training samples needed can be reduced by 4060 percent (Tuia *et al*., 2011). Monitoring well drilling campaigns should be targeted at high-uncertainty areas as determined using uncertainty mapping to help refine the model faster and keep a check on the costs of field campaigns.

3.9 Implications for Water Resource Management

The reported accuracies of random forest and gradient boosting models, *exceed commonly accepted thresholds for screening-level environmental prediction modeLS,* indicating that they are prepared to be used in operations in data-sparse areas. The explained variance of 85-87% in the explanations of nitrate predictions are comparable to mechanistic fate and transport models which usually find it difficult to explain beyond 70% on subsurface characterization calibration (Stigter *et al*., 2006). *The primary advantage of remote sensing-driven machine learning lies in its ability to provide continuous spatial coverage across large regions.*

The model outputs may be used to identify the areas of vulnerability that require increased monitoring, inform the location of new water supply wells to avoid areas with contamination, enable the best management practices to be implemented in the high-risk agricultural areas, in the case of groundwater protection planning. The temporal resolution of predictions monthly facilitates adaptive management since seasonal periods of risk of high contamination are identified and preventive actions should be prioritized during such periods. The early warning may be integrated to alert when precursor conditions (intense irrigation, the onset of drought) are observed by satellites in order to give time in advance to treat the wellhead or switch the source in case of any contamination event.

Climate scenario projections provide information on long-term planning of water resources under uncertainty in terms of decision making. The ensemble-based predictive probabilities allow riskaverse risk-taking which takes into consideration worst-case contamination paths instead of using individual deterministic forecasts. Projection ranges can be used by water utilities to estimate the adaptation requirements of infrastructure (e.g. increasing the capacity of the treatment infrastructure to cope with expected salinity or nitrate increases in high-emission publications). The Extension agricultural services may also decipher the contamination hotspots maps and tailor the outreach programs on nutrient management practices and irrigation efficiency enhancement.

Nonetheless, to convert model forecasts into management decisions, one should be cautious about conveying fundamental uncertainties as well as setting up ground-truting campaigns. The relative errors of 12-20% that are typical of current applications force sampling before expensive interventions e.g. well closures or the required land use limitations are realized. The deployment of operations is suggested to be carried out in a tiered fashion whereby remote sensing-based

screening is used to determine areas with potential candidates to be subjected to in-depth hydrogeological probing to guide high stakes decisions that would need more certainty.

4.0 Conclusion

This study demonstrates that the integration of machine learning models with multi-sensor remote sensing datasets provides a robust framework for predicting groundwater contamination across diverse hydrogeological environments. The results show that Random Forest and Gradient Boosting models achieved high predictive accuracy, with test-set R^2 values exceeding 0.80 for nitrate and above 0.75 for salinity, indicating their suitability for operational risk assessment and groundwater management applications.

Feature importance analysis revealed that remotely sensed variables—particularly vegetation indices, irrigation intensity, land surface temperature, and antecedent precipitation—serve as effective proxies for subsurface contamination processes. This confirms the strong conceptual linkage between surface environmental dynamics and groundwater quality. Furthermore, the incorporation of climate projections through data augmentation improved model generalization under extreme conditions and enabled ensemble-based forecasting of future contamination trends. Projected increases in nitrate concentrations (8–28%) across irrigated regions highlight the potential risks posed by climate change under moderate to high emission scenarios.

Despite these advances, model transferability across regions remains constrained by hydrogeological heterogeneity and region-specific environmental relationships. This limitation underscores the need for hybrid modeling approaches that integrate physical process understanding with data-driven techniques. In addition, uncertainty analysis identified key data gaps, particularly in areas with sparse monitoring networks and complex subsurface conditions, emphasizing the importance of targeted data acquisition and validation.

For practical implementation, the deployment of these models requires robust validation protocols and active stakeholder engagement to translate probabilistic predictions into actionable water resource management strategies. Overall, this study provides a scalable and adaptable framework for leveraging artificial intelligence and Earth observation data to support sustainable groundwater management under evolving climate conditions.

5.0 References

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). *Crop evapotranspiration: Guidelines for computing crop water requirements* (FAO Irrigation and Drainage Paper No. 56). Food and Agriculture Organization.

Anderson, M. C., Kustas, W. P., Norman, J. M., Hain, C. R., Mecikalski, J. R., Schultz, L., Gonzalez-Dugo, M. P., Cammalleri, C., d'Urso, G., Pimstein, A., & Gao, F. (2011). Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. *Hydrology and Earth System Sciences, 15*(1), 223–239. https://doi.org/10.5194/hess-15-223-2011

Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin, 24*(1), 43–69. https://doi.org/10.1080/02626667909491834

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324

Burow, K. R., Dubrovsky, N. M., & Shelton, J. L. (2007). Temporal trends in concentrations of DBCP and nitrate in groundwater in the eastern San Joaquin Valley, California, USA. *Hydrogeology Journal, 15*(5), 991–1007. https://doi.org/10.1007/s10040-006-0148-7

Cannon, A. J. (2018). Multivariate quantile mapping bias correction: An N-dimensional probability density function

transform for climate model simulations of multiple variables. *Climate Dynamics, 50*, 31–49. https://doi.org/10.1007/s00382-017-3580-6

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., & Meygret, A. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment, 120*, 25–36. https://doi.org/10.1016/j.rse.2011.11.026

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., & Kimball, J. (2010). The Soil Moisture Active Passive (SMAP) mission. *Proceedings of the IEEE, 98*, 704–716. https://doi.org/10.1109/JPROC.2010.2043918

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development, 9*, 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*, 1189–1232.

Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on Machine Learning* (pp. 1050–1059).

Gleeson, T., Cuthbert, M., Ferguson, G., & Perrone, D. (2020). Global groundwater sustainability, resources, and systems in the Anthropocene. *Annual Review of Earth and Planetary Sciences, 48*, 431–463. https://doi.org/10.1146/annurev-earth-071719-055251

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society, 90*, 1095–1108.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., & Simmons, A. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society, 146*, 1999–2049. https://doi.org/10.1002/qj.3803

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment, 25*, 295–309.

Jasechko, S., & Perrone, D. (2021). Global groundwater wells at risk of running dry. *Science, 372*, 418–421. https://doi.org/10.1126/science.abc2755

Jiang, S., Zheng, Y., & Solomatine, D. (2020). Improving AI system awareness of geoscience knowledge: Symbiotic integration of physical approaches and deep learning. *Geophysical Research Letters, 47*, e2020GL088229. https://doi.org/10.1029/2020GL088229

Kath, J., Reardon-Smith, K., Le Brocque, A. F., Apan, A., Mushtaq, S., & Stone, R. C. (2014). Groundwater decline and tree change in floodplain landscapes: Identifying non-linear threshold responses in canopy condition. *Global Ecology and Conservation, 14*, e00398.

Knutti, R., & Sedlaček, J. (2013). Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Climate Change, 3*, 369–373. https://doi.org/10.1038/nclimate1716

Kohler, M. A., & Linsley, R. K. (1951). *Predicting the runoff from storm rainfall*

(Weather Bureau Research Paper No. 34). U.S. Department of Commerce.

Lapworth, D. J., Nkhuwa, D. C. W., Okotto-Okotto, J., Pedley, S., Stuart, M. E., Tijani, M. N., & Wright, J. (2017). Urban groundwater quality in sub-Saharan Africa: Current status and implications for water security and public health. *Hydrogeology Journal, 25*, 1093–1116. https://doi.org/10.1007/s10040-016-1516-6

Leblanc, M. J., Tregoning, P., Ramillien, G., Tweed, S. O., & Fakes, A. (2009). Basin-scale, integrated observations of the early 21st century multiyear drought in southeast Australia. *Water Resources Research, 48*, W04519.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*, 18–22.

MacDonald, A. M., Bonsor, H. C., Taylor, R., Shamsudduha, M., Burgess, W. G., Ahmed, K. M., Mukherjee, A., Zahid, A., Lapworth, D., Gopal, K., & Rao, M. S. (2015). *Groundwater resources in the Indo-Gangetic Basin: Resilience to climate change and abstraction* (British Geological Survey Open Report OR/15/047). British Geological Survey.

Mukherjee, A., Fryar, A. E., Scanlon, B. R., Bhattacharya, P., & Bhattacharya, A. (2011). Elevated arsenic in deeper groundwater of the western Bengal basin, India: Extent and controls from regional to local scale. *Applied Geochemistry, 99*, 1–17.

Nolan, B. T., Gronberg, J. M., Faunt, C. C., Eberts, S. M., & Belitz, K. (2014). Modeling nitrate at domestic and public-supply well depths in the Central Valley, California. *Environmental Science & Technology, 48*, 5643–5651. https://doi.org/10. 1021/es405452q

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics, 378*, 686–707.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., & Warton, D. I. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography, 40*, 913–929. https://doi.org/10.1111/ecog.02881

Singh, S., Snehlata, Jaiswal, S., & Saxena, M. (2025, September). Feature-enriched environmental and temporal forecasting via enhanced XGBoost algorithm. In *International Conference on Next-Generation Networks and Deployable Artificial Intelligence* (pp. 446–458). Springer Nature Switzerland.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing, 14*, 199–222.

Stigter, T. Y., Ribeiro, L., & Dill, A. M. M. C. (2006). Evaluation of an intrinsic and a specific vulnerability assessment method in comparison with groundwater salinisation and nitrate contamination levels in two agricultural regions in the south of Portugal. *Hydrogeology Journal, 14*, 79–99. https://doi.org/10.1007/s10040-004-0396-3

Tartakovsky, A. M., Marrero, C. O., Perdikaris, P., Tartakovsky, G. D., & Barajas-Solano, D. (2020). Physics-informed deep neural networks for learning parameters and constitutive relationships in subsurface flow problems. *Water Resources Research, 56*, e2019WR026731. https://doi.org/10.1029/2019WR026731

Taylor, R. G., Scanlon, B., Doll, P., Rodell, M., van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J. S., Edmunds, M., & Konikow, L. (2013). Ground water and climate change. *Nature Climate Change, 3*, 322–329.

Themeßl, M. J., Gobiet, A., & Leuprecht, A. (2011). Empirical-statistical downscaling

and error correction of daily precipitation from regional climate models. *International Journal of Climatology, 31*(15), 1530–1544. https://doi.org/10.1002/joc.2168

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., & Traver, I. N. (2012). GMES Sentinel-1 mission. *Remote Sensing of Environment, 120*, 9–24. https://doi.org/10.1016/j.rse.2011.05.028

Tuia, D., Volpi, M., Copa, L., Kanevski, M., & Muñoz-Marí, J. (2011). A survey of active learning algorithms for supervised remote sensing image classification. *IEEE Journal of Selected Topics in Signal Processing, 5*, 606–617.

Vapnik, V. (1998). *The nature of statistical learning theory*. Springer.

Wan, Z., & Dozier, J. (1996). A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing, 34*, 892–905.

Wan, Z. (2008). New refinements and validation of the MODIS land-surface temperature/emissivity products. *Remote Sensing of Environment, 112*, 59–74.

World Health Organization. (2011). *Guidelines for drinking-water quality* (4th ed.). World Health Organization.

Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., Allen, R. G., Anderson, M. C., Belward, A. S., Cohen, W. B., & Dwyer, J. (2019). Current status of Landsat program, science, and applications. *Remote Sensing of Environment, 225*, 127–147. https://doi.org/10.1016/j.rse.2019.02.015

Zhou, B., Chao, Q., & Huang, L. (2015). The core conclusions and interpretation of working group I contribution to the fifth assessment report of the intergovernmental panel on climate change. *Chinese Journal of Urban and Environmental Studies, 3*(1), 1550003.

Zhu, C., & Burden, D. S. (2001). Mineralogical compositions of aquifer matrix as necessary initial conditions in reactive contaminant transport models. *Journal of Contaminant Hydrology, 51*, 145–161. https://doi.org/10.1016/S0169-7722(01)00132-2

Zhu, Z., & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment, 118*, 83–94. https://doi.org/10.1016/j.rse.2011.10.028

Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine, 5*, 8–36.

Declaration

Consent for publication

Not Applicable

Availability of data and materials

The publisher has the right to make the data public

Conflict of Interest

The authors declared no conflict of interest

Ethical Considerations

Not applicable

Funding

The author declared no source of funding

Authors' Contributions

MBA conceived the study, performed data analysis, and developed the machine learning framework. KK contributed to remote sensing data acquisition, preprocessing, and climate scenario integration. AO assisted in model validation, statistical evaluation, and interpretation of hydrogeological results. RE contributed to literature review, manuscript drafting, and critical revision of the intellectual content. All authors reviewed, approved, and agreed to the final version of the manuscript.