Ecotoxicology and Environmental Safety 323 (2026) 120627

Contents lists available at ScienceDirect

Ecotoxicology and Environmental Safety

journal homepage: www.elsevier.com/locate/ecoenv

Predicting and assessing the distribution of high-fluoride groundwater in China using multiple machine learning models: A big data-driven geospatial analysis

Mengyao Su [a,b,c], Qingbo Wang [a,b,c], Yunzhu Liu [a,b,c], Xin Wang [a,b,c], Chao Zhang [a,b,c], Xin Zhang [a,b,c], Chang Liu [a,b,c], Yanmei Yang [a,b,c], Peng Luo [c,d], Yue Gao [a,b,c,*], Yanhui Gao [a,b,c,*]

[a] Center for Endemic Disease Control, Chinese Center for Disease Control and Prevention, Harbin Medical University, Harbin 150086, PR China
[b] NHC Key Laboratory of Etiology and Epidemiology, Harbin Medical University, Harbin 150086, PR China
[c] Joint Key Laboratory of Endemic Diseases, Harbin Medical University, Guizhou Medical University, Xi'an Jiaotong University, PR China
[d] Guizhou Medical University, Guiyang 550025, PR China

ARTICLE INFO

Keywords:
Water fluoride
Random forest
Artificial neural network
Logistic regression
Environmental variables

ABSTRACT

Drinking-water fluorosis remains a major public health concern in China due to sustained reliance on high-fluoride groundwater in rural regions. Using the largest harmonized national dataset to date—138,180 site records and 38 environmental predictors spanning climatic, topographic, soil, hydrological, geological, and volcanic domains—artificial neural network (ANN), random forest (RF), and logistic regression (LR) models were developed and evaluated via spatial block 5-fold cross-validation to mitigate spatial autocorrelation and improve geographic generalization. All models demonstrated strong spatial generalizability, with precision-recall area under the curve (PR_AUC) values of 0.985 (ANN), 0.992 (RF), and 0.956 (LR). The resulting probabilistic risk maps revealed both well-established and previously under-recognized high-fluoride zones, including the Northeast Plain, North China Plain, Inner Mongolia, and parts of northwest China, as well as newly highlighted candidate areas such as the Jianghan Plain, Sanjiang Plain, and Junggar and Tarim basin rims. An estimated 40.78–54.77 million rural residents reside in predicted high-risk zones. Meteorological conditions, topographic–soil properties, and volcanic–geologic factors emerged as dominant controls on fluoride enrichment. This three-model comparative framework provides a validated, scalable tool for national groundwater fluorosis risk assessment and offers actionable insights to support targeted surveillance, mitigation strategies, and groundwater safety management, offering a framework applicable to environmentally similar regions such as Mongolia and Kazakhstan.

1. Introduction

Prolonged exposure to high-fluoride groundwater has caused endemic fluorosis, jeopardizing the health of millions globally (Karabulut et al., 2024; Nordstrom, 2022; Podgorski and Berg, 2022). China is among the nations most severely impacted by this issue, having identified affected areas in 28 provinces and the Xinjiang Production and Construction Corps, with 72.07 million individuals at risk (Yang et al., 2025). Water improvement initiatives have significantly elevated water quality for residents in recognized endemic regions; however, in certain areas, the naturally elevated fluoride concentrations in groundwater hinder the identification of appropriate low-fluoride sources, resulting in persistent areas where fluoride levels surpass established standards despite intervention (Nordstrom, 2022). A comprehensive understanding of the geological and hydrological elements is necessary to address this issue.

Groundwater fluoride concentrations are strongly shaped by environmental conditions. These concentrations depend on the distribution of fluoride-bearing minerals (Ali et al., 2026; Nawrin et al., 2024) and are affected by geomorphology, hydroclimate, soil properties, volcanic

* Correspondence to: Center for Endemic Disease Control, Chinese Center for Disease Control and Prevention, Harbin Medical University, No.157 Baojian Road, Nangang District, Harbin 150081, PR China.

E-mail addresses: gao15663673888@163.com (Y. Gao), gaoyanhui@hrbmu.edu.cn (Y. Gao).

https://doi.org/10.1016/j.ecoenv.2026.120627
Received 7 September 2025; Received in revised form 1 August 2026; Accepted 4 August 2026
Available online 6 August 2026
0147-6513/© 2026 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

activity, and human activities (Abiye et al., 2018; Durrani and Farooqi, 2021; Jia et al., 2018; Liu et al., 2022; Rubio et al., 2020; Senarathne et al., 2022; Solanki et al., 2022; Yuan et al., 2023). Previous studies have established that the dissolution of fluorine-rich rocks in aquifers and the presence of fluorides in soil are considered the primary sources of groundwater fluoride (Nakaya et al., 2023; Zhao et al., 2023). Furthermore, soil physicochemical properties affect fluoride migration and accumulation in groundwater (Subbaiah et al., 2023; Zhao et al., 2023; Zhu et al., 2025). For example, alkaline soils, high cation exchange capacity, and low organic matter content promote the transfer of fluoride from soil to groundwater (Alcaine et al., 2020; Kazapoe et al., 2024; Zhang et al., 2025). Climate also plays an important role by influencing groundwater recharge and storage; dry and semi-arid regions are generally more favorable for fluoride enrichment (Wang et al., 2022). Volcanic eruptions may further increase fluoride levels by releasing ash that contaminates soils and enhances fluoride transport to groundwater (Virchow et al., 2026). Human activities, including agricultural irrigation and industrial production, can also contribute to fluoride accumulation (Huang et al., 2022; Li et al., 2021; Wang et al., 2019; Xu et al., 2022). Therefore, a systematic evaluation of these environmental factors is essential for predicting the national-scale spatial distribution of groundwater fluoride.

In response to this need, machine learning methods have been increasingly applied to map geogenic groundwater contaminants at regional to global scales, leveraging their strength in handling nonlinear relationships within large multivariate datasets (Barzegar et al., 2017; Gupta and Maiti, 2022; Ling et al., 2022; Lu et al., 2023; Podgorski et al., 2018; Rosecrans et al., 2022; Sarkar et al., 2022; Yang et al., 2024). Podgorski and Berg (2022) employed a random forest (RF) model to generate a global groundwater fluoride hazard map, estimating that approximately 180 million people are at risk. Mohammadi et al. (2016) utilized an artificial neural network (ANN) to forecast fluoride distribution in Khaf City, Iran, supporting local fluorosis prevention. Nafouanti et al. (2021) combined ANN, RF, and binary logistic regression (LR) in the Datong Basin, China, demonstrating multi-model value. These studies illustrate the potential of data-driven approaches to identify high-risk zones beyond traditional sampling networks.

However, most national-scale assessments in China remain limited by relatively small sample sizes, incomplete environmental covariates, and reliance on single-model frameworks that preclude assessment of prediction uncertainty and model-specific bias, limiting reliability for public health decision-making (Cao et al., 2022). Conventional risk maps derived from historical endemic village records inherently underrepresent unmonitored areas, lack spatially explicit probabilistic assessment, and cannot reveal hidden high-risk zones. Moreover, standard random cross-validation often overlooks spatial autocorrelation, leading to over-optimistic performance estimates and reduced real-world generalizability. Consequently, thousands of rural communities in geologically suspect but unmonitored regions may remain exposed to unsafe fluoride levels without being prioritized for water improvement interventions, underscoring the urgent need for more comprehensive, spatially validated risk mapping.

To address these deficiencies, this study established the first multi-model comparative framework for the spatially cross-validated prediction of high-fluoride groundwater risk in China. This framework was implemented through three objectives: (i) to harmonize the largest national geogenic fluoride database to date—comprising 138,180 site records (70,479 high-fluoride measurements and 67,701 control sites) across 28 provinces—and identify key environmental predictors of fluoride enrichment across diverse climatic and geologic settings in China, drawing on 38 predictors spanning climatic, topographic, soil, hydrological, geological, and volcanic domains; (ii) to evaluate the complementary predictive performance and spatial generalizability of ANN, RF, and LR under spatial block cross-validation; and (iii) to generate spatially explicit probabilistic risk maps that identify both established endemic zones and previously unrecognized candidate high-risk areas, and to quantify the rural population in predicted high-risk areas, thereby informing evidence-based prioritization of surveillance and optimization of water-source selection. This strategy offers a replicable framework for environmentally similar fluoride-affected regions.

2. Materials and methods

2.1. Data sources and preprocessing

2.1.1. Data sources

All fluoride data were collected based on the 1985 national village sampling framework for fluorosis and water-improvement surveys (Wang et al., 2012). These records represent a single-phase cross-sectional survey of fluoride concentrations in drinking-water sources of endemic villages. Post-intervention data were excluded because extensive water-improvement projects altered natural geogenic fluoride distributions. The final dataset captures pre-intervention baseline conditions across 28 provincial-level administrative regions in China. High-fluoride sites correspond to historically documented endemic fluorosis areas identified in prior national surveys. Control sites were generated at a 1:1 ratio outside 10-km buffers around high-fluoride localities using random sampling in ArcGIS 10.0.

A total of 38 environmental variables were collected from open-access geospatial datasets, covering climate, topography, soil, hydrology, geology, and volcanic activity (Table S1). Spatial processing was conducted in ArcGIS 10.0 to derive variables including river density, distance to rivers, crater density, distance to volcanic centers, slope, aspect, and topographic wetness index.

2.1.2. Data preprocessing and georeferencing

Data cleaning was performed to remove duplicates, missing values, and anomalous records. After quality control, 70,479 high-fluoride records and 67,701 control records were retained, yielding a total of 138,180 site records. Fluoride concentrations were measured using the ion-selective electrode method following GB/T 5750–2006, with strict quality control applied at county and municipal Centers for Disease Control and Prevention. Village-level coordinates in the BD−09 datum were obtained using the Baidu Maps Geocoding API (http://api.map.baidu.com/geocoding/v3/), with a numerical precision of 10^{-6} degrees. The geocoding accuracy is shown in Figure S1. These coordinates were subsequently converted to WGS 1984 geographic coordinates using a coordinate transformation algorithm (BD−09 to WGS 1984). Control sites were directly generated in the WGS 1984 system. All sampling locations were stored as point features in ArcGIS 10.0 for spatial analysis (Fig. 1).

2.1.3. Variable preprocessing

All predictor variables were standardized before modeling. Continuous variables were rescaled to the range [0, 1] to eliminate scale effects. Three categorical variables—geological age, lithology, and soil unit symbol—were converted into dummy variables. Univariate logistic regression was used to screen preliminary associations between each variable and high-fluoride occurrence.

Multicollinearity was assessed using variance inflation factors (VIF). Tables S2−4 present the hierarchical classifications of geological lithology (Geo_lithology), geological age (Geo_age), and soil unit symbol (Su_sym), respectively.

2.2. Feature selection

Feature selection was tailored to each model: for LR, stepwise selection was applied within each cross-validation fold ($\alpha\_in = 0.05$, $\alpha\_out = 0.10$), retaining 17 predictors; for RF, variable importance was ranked by mean out-of-bag (OOB) permutation and variables with importance $> 1\%$ were retained, yielding 16 predictors; and for ANN, Recursive Feature Elimination (RFE) retained the top 18 predictors across all folds,

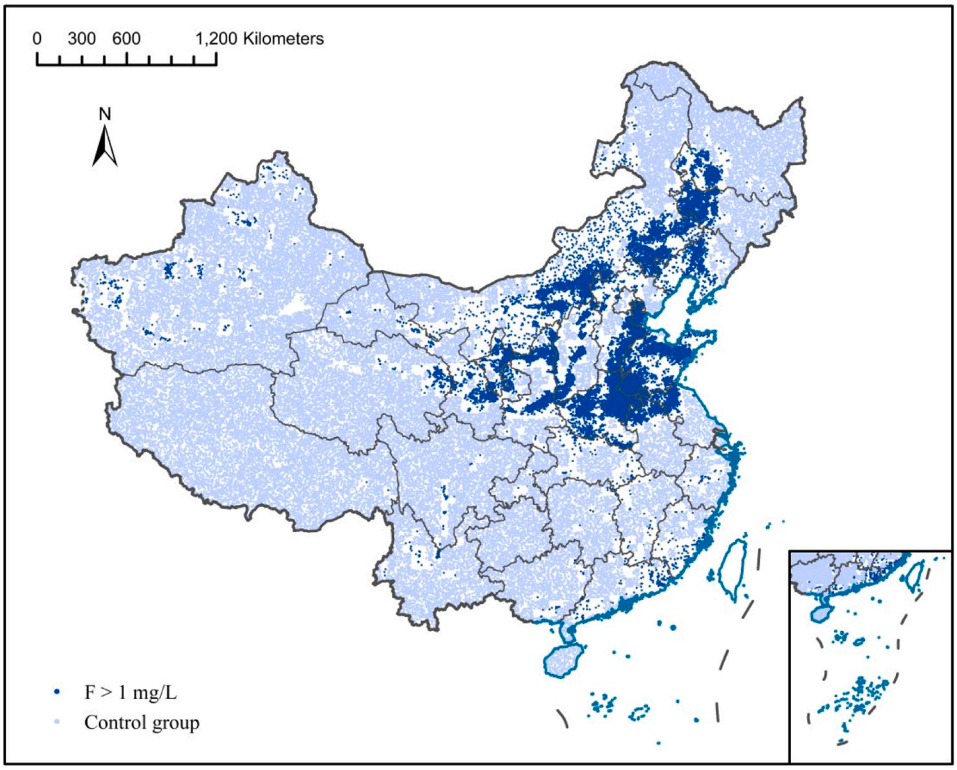

Fig. 1. Distribution of High-Fluoride Groundwater and Control Groups.

with L2 regularization and early stopping controlling model complexity. The network architecture of the ANN model is shown in Table S5.

2.3. Cross-validation strategy

Two cross-validation strategies were used to evaluate model performance: random 5-fold cross-validation (R-CV) and spatial block 5-fold cross-validation (S-CV). For R-CV, all samples were randomly shuffled and partitioned into 5 approximately equal folds; in each iteration, the *i*-th fold ($i = 1$–5) was used as the validation set and the remaining four folds were combined as the training set. For S-CV, China was divided into 50 km $\times$ 50 km grid blocks (n = 3982). Blocks were assigned to five folds using spatial block sampling with a contiguous grouping strategy, ensuring that training and validation data were drawn from distinct geographic regions, thereby reducing the influence of spatial autocorrelation.

2.4. Model training

Within each fold, candidate hyperparameter configurations were tuned solely on the training partition, and the configuration achieving the highest internal validation AUC was assessed on the independent validation blocks. Final models were retrained on the full dataset (n = 138,180) using the selected hyperparameters.

2.5. Model evaluation

Model performance was assessed using accuracy, sensitivity, specificity, precision, F1-score, the area under the receiver operating characteristic curve (AUC), and the area under the precision-recall curve (PR_AUC). Prediction uncertainty was evaluated using predicted probabilities: values near 0.5 indicate high uncertainty, while values close to 0 or 1 indicate high confidence. Spatial agreement between predicted high-risk zones and observed fluoride distributions was evaluated to ensure realistic geographic patterns. The definitions and computational methods for these metrics are provided below:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{1}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{5}$$

$$\text{AUC} = \int_0^1 \text{Sensitivity}(1 - \text{Specificity})\text{d}(1 - \text{Specificity}) \tag{6}$$

$$\text{PR\_AUC} = \int_0^1 \text{Precision}(\text{Sensitivity})\text{d}(\text{Sensitivity}) \tag{7}$$

Here, TP represents the number of true positives, referring to high-fluoride groundwater samples correctly classified as high-fluoride. TN represents the number of true negatives, i.e., non-high-fluoride samples correctly classified as non-high-fluoride. FP represents the number of false positives, referring to non-high fluoride samples incorrectly classified as high-fluoride. FN represents the number of false negatives, referring to high-fluoride samples incorrectly classified as non-high-fluoride.

2.6. Generation of prediction maps

National prediction grids were created at 0.05° $\times$ 0.05° resolution using the three trained models. Continuous probability surfaces were generated using inverse distance weighting (IDW) in ArcGIS 10.0

(power = 2, 12 nearest neighbors) to ensure spatial smoothness and avoid mosaic effects. High-risk zones were defined using a probability threshold of ≥ 0.5. IDW was applied post-prediction solely for cartographic smoothing; all statistical evaluations were based on raw model outputs.

2.7. Population risk assessment

Rural population data for 2020 at 1 km × 1 km resolution were obtained from the Resource and Environmental Science and Data Center (RESDC; https://www.resdc.cn/). Urban areas were masked to retain only rural populations. The rural population residing in predicted high-risk areas (probability ≥ 0.5) was quantified by overlaying the population grid with the prediction surface, representing an upper-bound estimate given the absence of groundwater-source-specific population data. Sensitivity analysis was performed using thresholds of 0.4, 0.5, and 0.6 to evaluate the robustness of these population estimates.

3. Results

3.1. Analysis of key predictive indicators

The detailed coefficients, statistical significance, and effect sizes (odds ratios) for all predictors retained in the final logistic regression model are presented in Table S7. All continuous variables were normalized to a [0,1] range. The results demonstrate that soil base saturation (T_bs) had the strongest positive association with high-fluoride risk (OR = 27.90, 95% CI: 24.770–31.452), while elevation exhibited the strongest negative association (OR = 0.004, 95% CI: 0.003–0.004). With the exception of the constant, all variables were statistically significant ($p < 0.001$).

3.2. Environmental controls on groundwater fluoride enrichment

All 38 environmental variables showed significant associations with high-fluoride groundwater occurrence in univariate logistic regression analyses ($p < 0.001$; Table S6). Variables with VIF > 5 were excluded, including topsoil sand fraction (T_sand), topsoil clay fraction (T_clay), reference bulk density of topsoil (T_ref_bulk_density), pressure (PRS), relative humidity (RHU), and sunshine days (SSD). The predictor variables retained in each final model are detailed in Table S8, and their national-scale spatial distributions are shown in Figure S2.

Variable importance differed markedly among the three models (Fig. 2), reflecting distinct sensitivities to environmental controls on groundwater fluoride enrichment. The ANN model primarily responded to dynamic geochemical and soil-salinity parameters, with elevation, topsoil salinity (T_ece), the cation exchange capacity of the clay (T_cec_clay), Haplic Kastanozem (Su_sym_KSh) and evaporation (EVP) identified as dominant predictors (Fig. 2A). In contrast, the RF model showed greater sensitivity to spatial configuration and regional geologic context, particularly proximity to volcanic centers (Volc_dist), alongside evaporative influences (Fig. 2B). The LR model, by comparison, relied more heavily on categorical geological and soil attributes, including Calcic Fluvisols (Su_sym_FLc), Quaternary age (Geo_age_Q), elevation, and soil properties such as T_bs (Fig. 2C). Nonlinear analyses revealed elevated fluoride probabilities in low-altitude plains (<200 m; Figure S3A), regions with annual evaporation of 1500–2000 mm (Figure S3B), and moderately wet landscapes with topographic wetness index (TWI) 15–20 (Figure S3C). Soil conditions including T_cec_clay = 50–70%, T_ece = 0–2 dS/m, T_bs > 90%, Topsoil Sodicity (T_esp) = 2–4%, and topsoil calcium carbonate content (T_CaCO_3) = 10–20% promoted fluoride mobilization (Figure S3E, F, G, H, M). Fluoride probability peaked at 30–35% within 200 km of volcanic centers

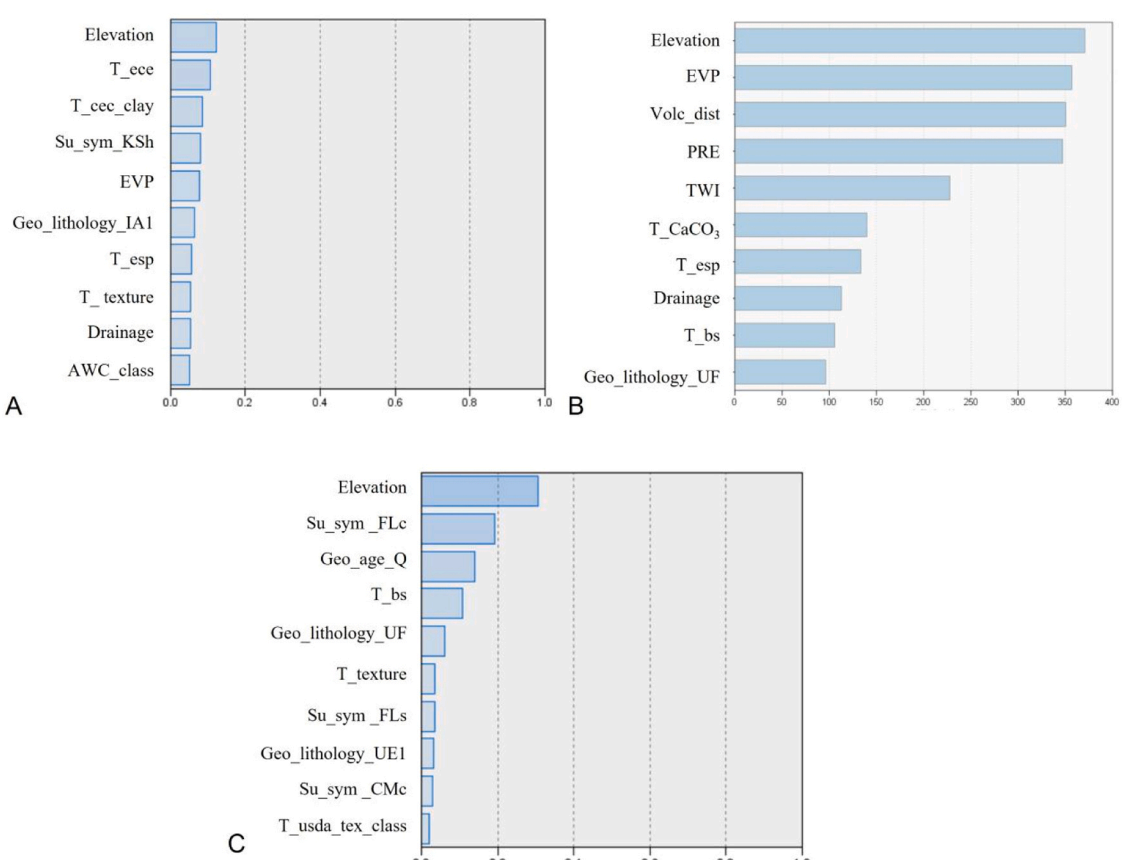

Fig. 2. Importance of the top ten environmental predictor variables in the models. (A) The top ten important variables in the ANN model, (B) The top ten important variables in the RF model, (C) The top ten important variables in the LR model.

(Figure S3D) and decreased gradually beyond 400 km (Figure S3D).

3.3. Groundwater fluoride risk prediction map

Cross-validation results for the ANN, RF, and LR models (Fig. 3) revealed clear differences between random 5-fold cross-validation (R-CV) and spatial block 5-fold cross-validation (S-CV). Under R-CV, all models showed high and stable performance, with mean AUCs of 0.992 $\pm$ 0.001 (ANN), 0.990 $\pm$ 0.001 (RF), and 0.961 $\pm$ 0.002 (LR). In contrast, S-CV led to notably lower mean AUCs and substantially increased variability (SD increased by approximately 6–20-fold), yielding AUCs of 0.972 $\pm$ 0.011, 0.980 $\pm$ 0.007, and 0.943 $\pm$ 0.024, respectively. These discrepancies indicate that R-CV likely underestimates prediction uncertainty by neglecting spatial autocorrelation, whereas S-CV provides a more realistic assessment of model generalizability. Additional performance metrics under S-CV (sensitivity, specificity, accuracy, F1-score, and PR_AUC) further demonstrated that the RF model exhibited the greatest stability across folds, while ANN showed moderate variability and LR the largest fluctuations, although all models maintained excellent discrimination ability (PR_AUC > 0.95).

3.4. Spatial distribution of high-fluoride groundwater

Based on the ANN, RF, and LR models, probabilistic maps of high-fluoride groundwater were generated across China. Areas with contamination probabilities > 0.5 accounted for approximately 25.2%, 20.0%, and 21.0% of the national territory, respectively (Table S9). High-risk zones were primarily concentrated in central, eastern, and northeastern China. The spatial overlap among the three models covered 13.7% of the country, including the Northeast Plain, North China Plain, Inner Mongolia, and parts of northwest China (Fig. 4D). Although the models agreed on core risk areas, distinct spatial patterns emerged due to differing predictor sensitivity. ANN highlighted zones governed by soil geochemistry and salinity gradients, RF emphasized regions influenced by volcanic proximity and evaporation, and LR delineated broad low-lying zones defined by geological and soil type. Beyond previously documented endemic regions, this study identified new candidate high-risk zones, including the Jianghan Plain, Sanjiang Plain, and Junggar and Tarim basin rims.

3.5. Prediction reliability and uncertainty

To assess the reliability of the spatial predictions, we analyzed the distribution of prediction probabilities (Fig. 5; Table S10). High-

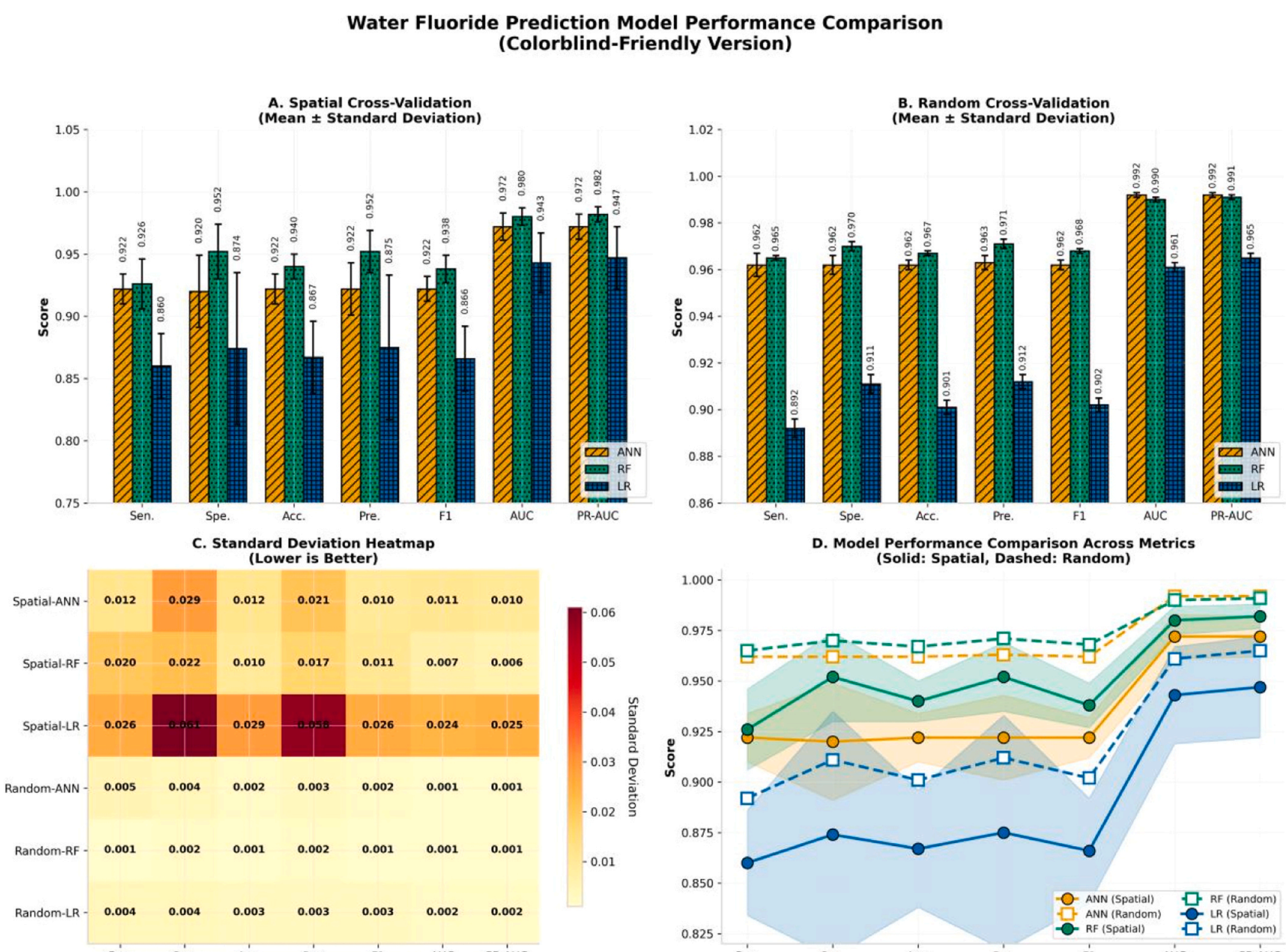

Fig. 3. Performance comparison of machine learning models for water fluoride prediction under spatial and random cross-validation strategies. (A) Spatial cross-validation results showing mean performance with standard deviation error bars. (B) Random cross-validation results exhibiting lower variability. (C) Heatmap of standard deviations across models and metrics (lower values indicate better stability). (D) Performance trend lines comparing spatial (solid) and random (dashed) validation across all metrics. ANN, artificial neural network; RF, random forest; LR, logistic regression; Sen., sensitivity; Spe., specificity; Acc., accuracy; Pre., precision; AUC, area under the ROC curve; PR_AUC, area under the precision-recall curve.

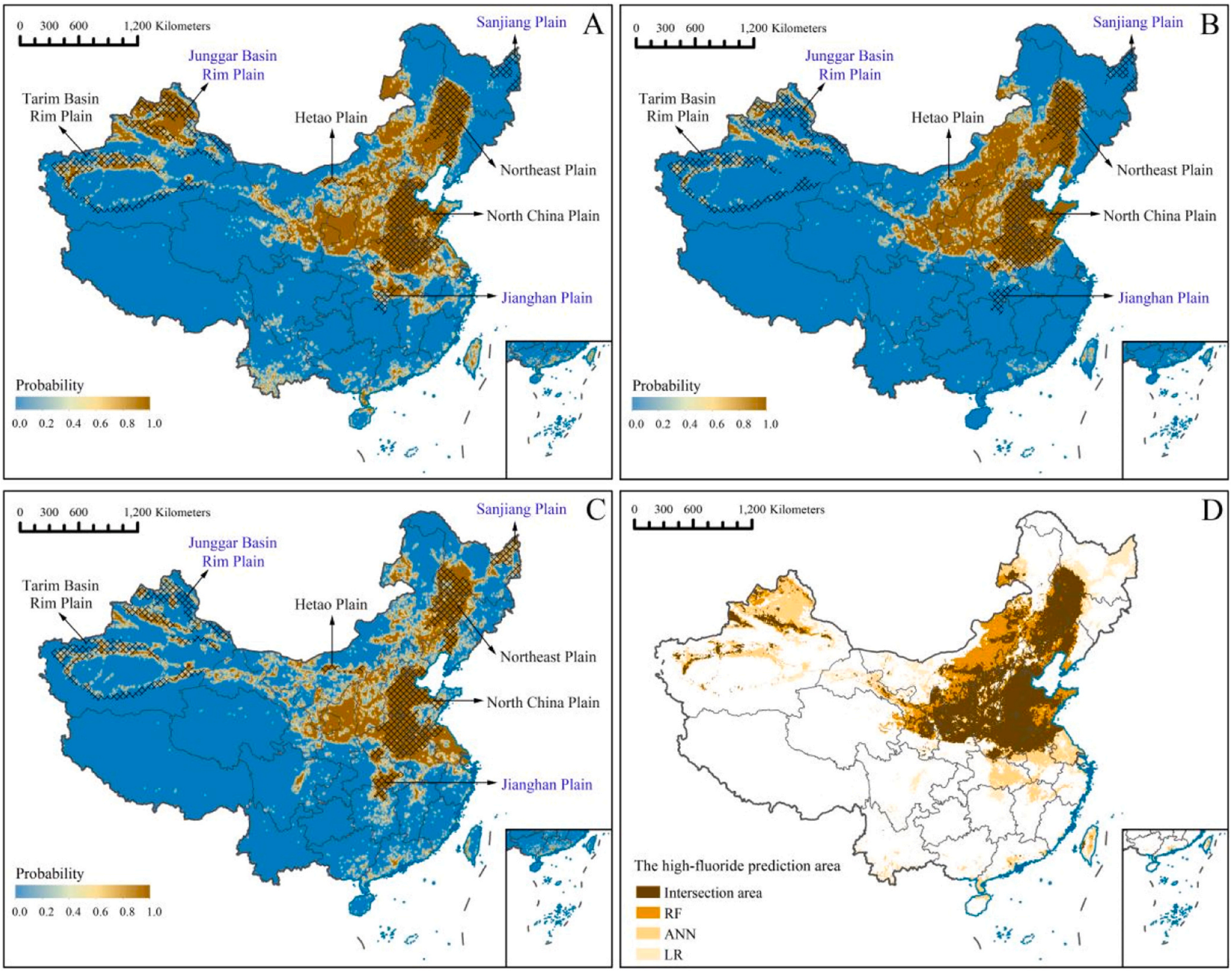

Fig. 4. Probability of groundwater fluoride exceeding 1 mg/L in predictive modeling. (A) The spatial range predicted by ANN modeling, (B) The spatial range predicted by RF modeling, (C) The spatial range predicted by LR modeling, (D) The spatial range and intersection range of high-fluoride risk areas predicted by ANN, RF, and LR models. Notes: Colors closer to blue indicate a stronger tendency toward low-fluoride areas, whereas colors closer to brown indicate a stronger tendency toward high-fluoride areas.

confidence high-fluoride regions were delineated as grid cells with projected probability ranging from [0.6–1.0), low-confidence (presumably safe) areas as [0–0.4), and ambiguity zones as [0.4–0.6). High-confidence high-risk zones accounted for 15.9–21.5% of China and were concentrated in historically endemic regions. Low-confidence safe zones occupied 72.3–78.5%, mainly in southern and western China. The uncertainty zones [0.4–0.6) accounted for 6.2%, 4.2%, and 10.0% of the country for ANN, RF, and LR, respectively, predominantly located in the transitional boundaries between high- and low-fluoride regions. ANN had the smallest uncertainty region and the highest percentage of high-confidence predictions, indicating that it offered the most stable and conservative spatial classification among the three models.

3.6. Performance of prediction models and accuracy of prediction maps

Compared with previous groundwater fluoride prediction studies, the models developed in this work exhibited markedly higher predictive performance and map accuracy. In the final model evaluation, the ANN, RF, and LR models achieved PR_AUC values of 0.985, 0.992, and 0.956, respectively (Fig. 6), with overall accuracies $\geq$ 87%. Agreement between predicted high-fluoride zones and documented high-fluoride villages was high, with village-level match rates ranging from 87.0% to 95.3% (Table S11). Calibration scatter plots indicated acceptable overall calibration for all three models, particularly at higher probabilities (>0.6), though systematic biases were evident in the low-probability range. The ANN model tracked the 1:1 line most consistently, while the LR model tended to underestimate risk below 0.4 predicted probability, and the RF model showed slight overestimation between 0.1–0.3 (Figure S4).

In contrast, previously reported national- or regional-scale machine-learning models generally showed lower discriminatory performance. As summarized in Table S12, earlier ANN-based models for China reported AUC values around 0.86 and accuracies near 80%, while RF-based models applied at national, continental, or regional scales in China and other regions typically achieved AUC values of approximately 0.85–0.92 or R^2 values between 0.5 and 0.9. Relative to these benchmarks, all three models in the present study demonstrated improved discrimination between high- and low-fluoride groundwater and higher spatial prediction accuracy at the national scale.

3.7. Estimation of population at risk in high-fluoride prediction zones

By overlaying predicted high-risk areas (probability $\geq$ 0.5) with rural population distribution, we estimated that approximately 51.15 $\pm$ 2.32

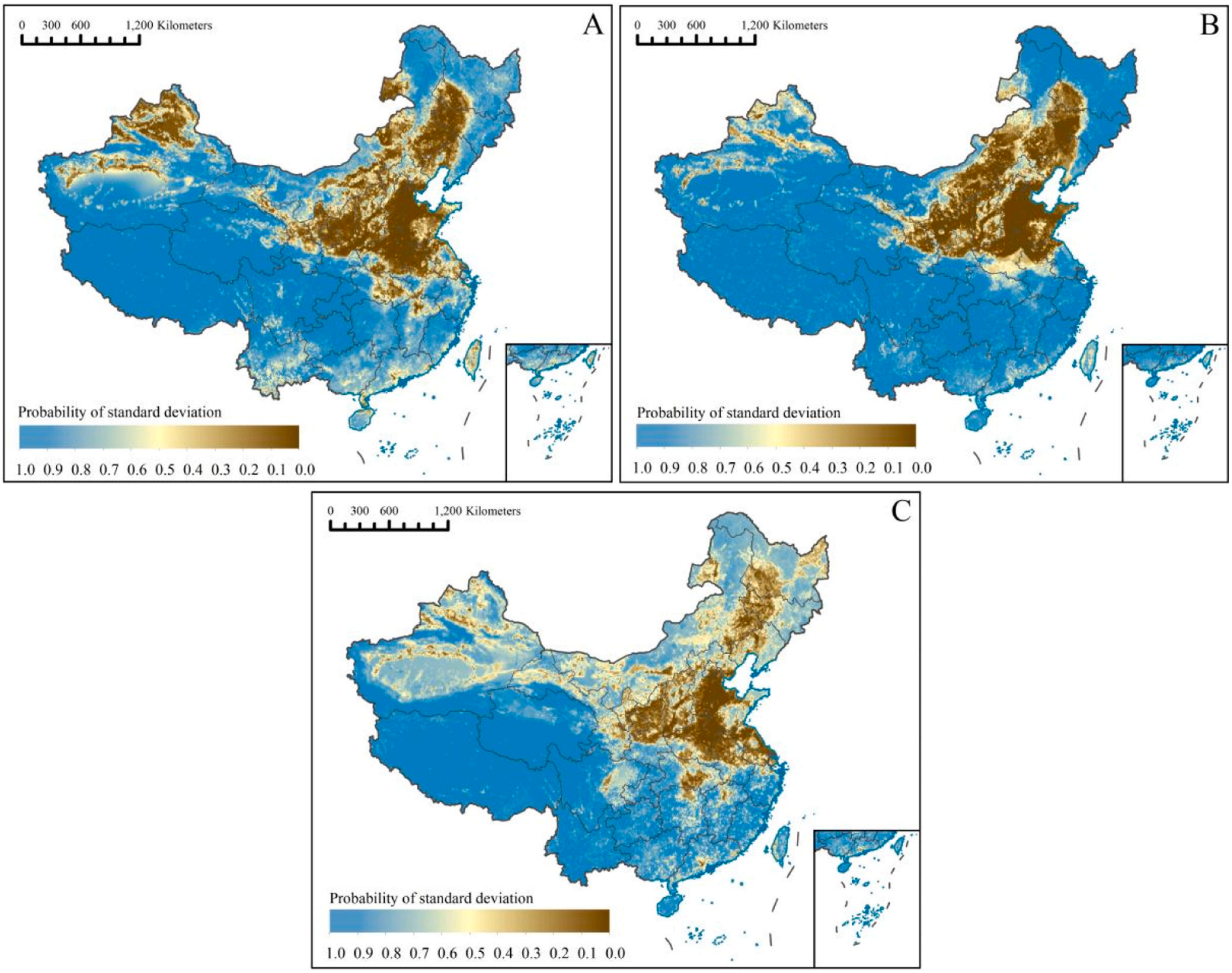

Fig. 5. Confidence maps for predicted groundwater fluoride levels in the models. (A):ANN modeling, (B):RF modeling, (C):LR modeling. Notes: The closer the color is to blue, the higher the predicted probability of low-fluoride areas; the closer the color is to brown, the higher the predicted probability of high-fluoride areas. Colors approaching yellow indicate lower prediction accuracy.

million, 40.78 ± 0.77 million, and 54.77 ± 1.24 million rural residents resided in these zones according to the ANN, RF, and LR models, respectively (Figure S5). The population in predicted high-risk areas showed pronounced spatial clustering, with the North China Plain representing the region with the highest density of rural residents within predicted high-risk areas (Fig. 7; Table S13). Additional population concentrations were evident in Shaanxi, Xinjiang, Inner Mongolia, Shanxi, and Gansu. Sensitivity analyses using alternative probability thresholds (0.4, 0.5, and 0.6) yielded a monotonic decline in estimated population size across all models (Figure S5). Among the three approaches, the RF model demonstrated the greatest robustness, exhibiting the smallest variability across thresholds, whereas ANN showed the largest fluctuations and LR exhibited intermediate stability. Across the full threshold range, the estimated population in predicted high-risk areas ranged from approximately 39.07–58.90 million. Together, these results quantified the rural population residing in predicted high-risk areas and identified priority regions where continued surveillance and field verification should be strengthened.

4. Discussion

4.1. Environmental controls on fluoride enrichment

This study advances national-scale prediction by integrating ANN, RF, and LR within a harmonized database of 138,180 records. Elevation (<200 m) consistently ranked first across all models, reflecting topographic control on groundwater residence time and drainage efficiency in low-lying plains. EVP (1500–2000 mm) featured prominently in ANN and RF, indicating evaporative concentration in semi-arid settings (Fig. 2; Figure S3A, B).

Low-lying plains accumulate fluoride through prolonged residence times and weak drainage, whereas arid regions concentrate solutes via evapotranspiration (Nakaya et al., 2023; Nordstrom, 2022; Wang et al., 2022). LR identified fine-grained Quaternary deposits (Geo_age_Q, Geo_lithology_UF) as favorable settings for fluoride retention, with Su_sym_FLc soils additionally marking alkaline, high-base-saturation environments (Fig. 2). Beyond these linear associations, ANN and RF captured complementary process-level mechanisms: ANN highlighted evaporative concentration and ion exchange via EVP, T_ece, and T_cec_clay (50–70%) (Figure S3B, E, G). Elevated temperature and evapotranspiration enhance solute enrichment in low-altitude plains, while high cation exchange capacity in clay minerals facilitates fluoride

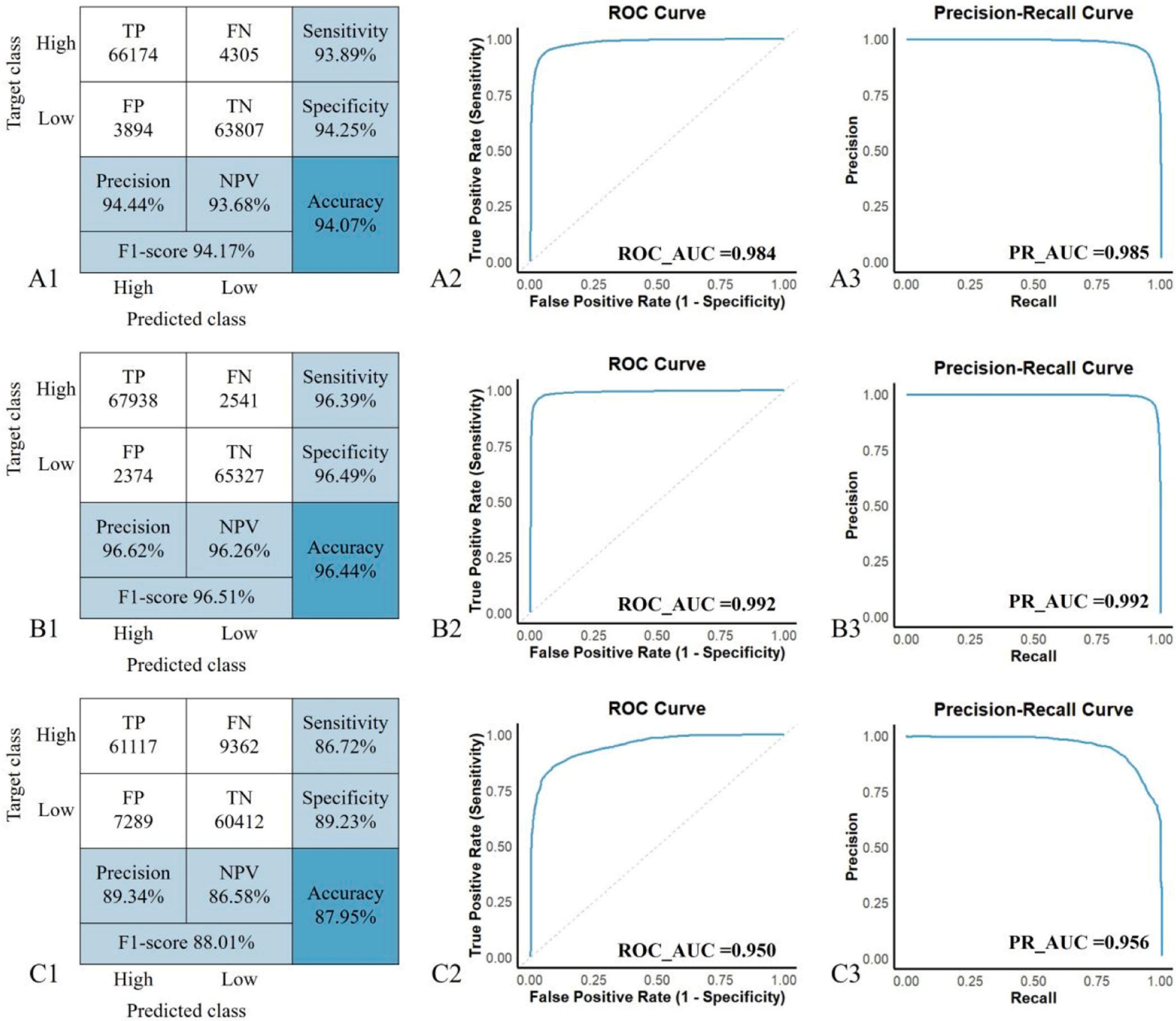

Fig. 6. Performance evaluation of the final ANN (A), RF (B), and LR (C) models based on the complete dataset. (A1, B1, C1) Confusion matrices (probability cutoff = 0.5). (A2, B2, C2) ROC curves (AUC). (A3, B3, C3) Precision-Recall curves (PR_AUC).

adsorption–desorption dynamics. RF separately highlighted proximity to volcanic sources (Volc_dist < 400 km) as a dominant spatial control, capturing discrete hotspots generated by magmatic degassing or hydrothermal fluids (Nordstrom, 2022; Figure S3D). Additionally, RF further corroborated the calcium-alkalinity pathway through T_$CaCO_3$ (10–20%) and T_esp (2–4%) across nonlinear climate-soil interactions (Figure S3F, M).

Collectively, these convergent findings—spanning topographic–hydroclimate consensus (elevation <200 m, EVP 1500–2000 mm across all models), linear hydrogeological templates (LR), nonlinear evaporative and soil-chemical controls (ANN), and volcanic–geochemical dual controls (RF)—validate model robustness by demonstrating that fluoride enrichment is governed by multiple, independently detectable hydrogeochemical pathways. The differential sensitivities of the three models thus provide a more robust and mechanistic understanding of national-scale fluoride controls than any single algorithm. Given environmental similarities with other affected regions (e.g., Mongolia, Kazakhstan), these mechanisms may inform global fluoride risk assessments.

4.2. Spatial convergence and divergence in predicted high-risk areas

Geographically, our forecasts correspond closely with historically documented high-fluoride regions—the Songnen Plain, North China Plain, Hetao Region, Loess Plateau, and Xinjiang (Li et al., 2019; Sun et al., 2026; Zhao et al., 2023)—while extending beyond traditionally recognized endemic villages along North China peripheries, central Shaanxi, and selected areas of Inner Mongolia and Xinjiang (He et al., 2020). New candidate regions identified by individual models include the Jianghan Plain, Sanjiang Plain, and Junggar and Tarim basin rims (Fig. 4). These peripheral and candidate areas share key environmental characteristics with established high-fluoride regions: low relief and evaporative climatic conditions. The geographic partitioning reveals limited core overlap (13.7%, Table S9) alongside extensive model-specific extensions. This pattern does not indicate a single-factor control mechanism; rather, it reflects differential algorithmic sensitivity to multiple, coexisting hydrogeochemical drivers.

The spatial disagreements shown in Fig. 4 reflect divergent model sensitivities to distinct fluoride controls. ANN, driven by T_ece and T_cec_clay, extended high-risk zones into the Jianghan Plain and the

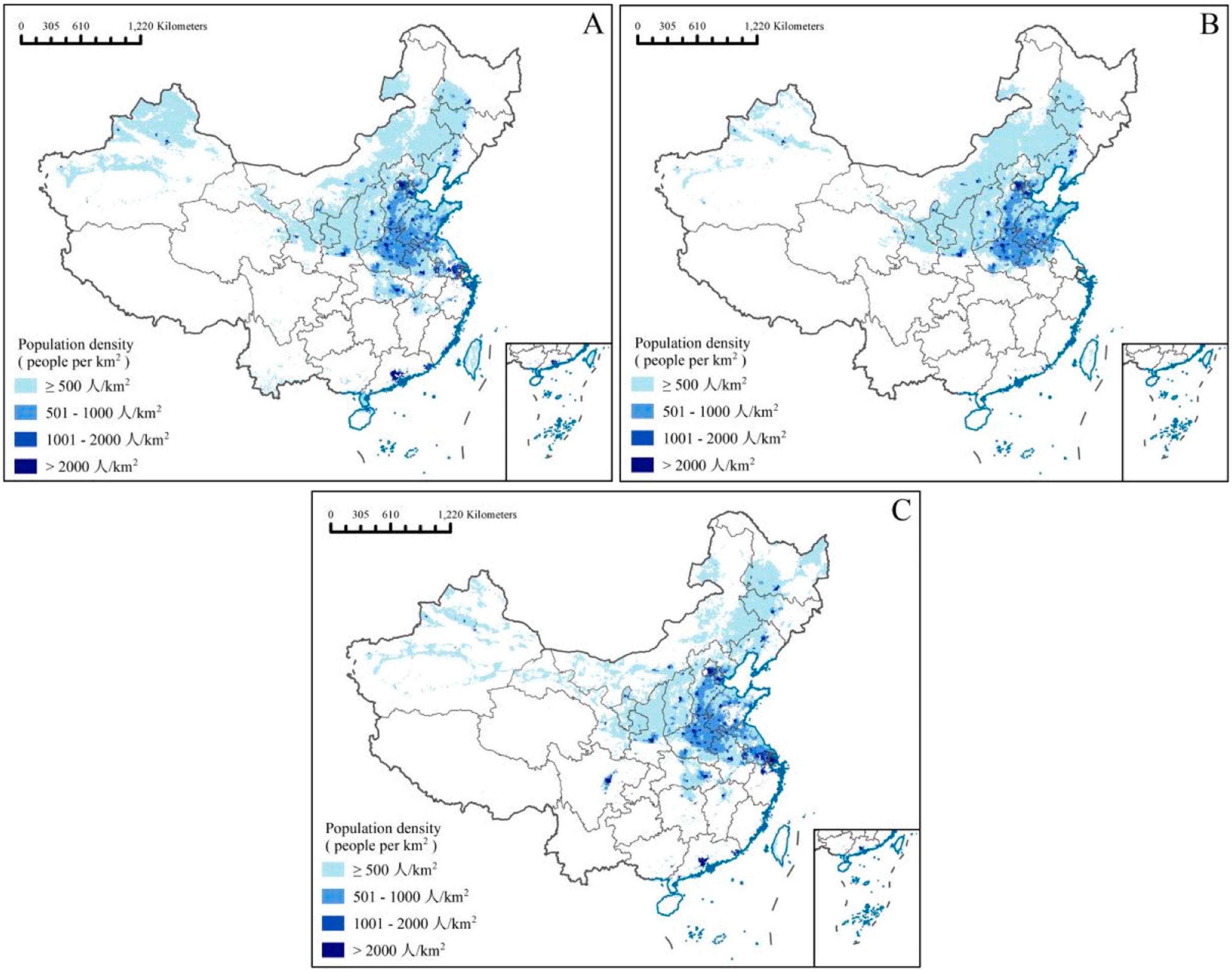

Fig. 7. Population density in predicted high-risk areas (predicted probability $\geq$ 0.5). ANN modeling (A), RF modeling (B), LR modeling (C). Notes: Lighter blue indicates lower population density within predicted high-risk zones; darker blue indicates higher density.

southern transitional fringes of the North China Plain, capturing localized salt-affected micro-domains (Fig. 4A). RF, giving heavy weight to Volc_dist, precipitation (PRE), and TWI, expanded coverage along the Hetao–Inner Mongolia transition, where its spatial-configuration metrics captured evaporative enrichment and drainage conditions beyond the alluvial-core template (Fig. 4B). LR, anchored to categorical Quaternary age, unconsolidated fluvial deposits and soil types (Geo_age_Q, Geo_lithology_UF, Su_sym_FLc), blanketed the Sanjiang Plain and broad low-lying plains (North China) with extensive, lower-probability zones, yet its predictions attenuated at basin margins and piedmont transitions where lithology deviates from the loose-sediment template (Fig. 4C). Collectively, these complementary perspectives demonstrate that no single algorithm captures the full spectrum of fluoride controls. Their consensus identifies regions where multiple controls converge, while their divergence marks zones where individual mechanisms dominate. Recent field investigations reporting elevated fluoride levels outside conventionally defined high-fluoride villages further support the effectiveness of model-guided exploration (Zhao et al., 2023). Together, these findings underscore the value of continuous, spatially explicit risk mapping for revealing the true extent of fluoride contamination and guiding targeted field validation.

4.3. *Model performance and spatially structured uncertainty*

Compared with previous national-scale (AUC $\approx$ 0.86) or global (AUC $\approx$ 0.90) assessments, the present national-scale three-model comparative framework demonstrated substantially improved predictive performance and broader applicability (Cao et al., 2022; Podgorski and Berg, 2022). Under both random and spatial cross-validation, all three models achieved high predictive accuracy, with AUC and PR_AUC values exceeding 0.95. Among them, the RF model exhibited the greatest stability across folds, with minimal variability in performance metrics, indicating a superior capacity to capture nonlinear relationships and spatial heterogeneity. The ANN model showed moderate variability, whereas the LR model displayed the largest fluctuations, likely reflecting its limited ability to represent complex interactions among environmental variables. These differences underscore the importance of employing multiple modeling approaches to enhance robustness and reduce algorithm-specific bias.

The spatial variation in predicted fluoride probability exhibited three distinct patterns. First, within core high-risk plains (North China, Northeast Plain), probabilities remained uniformly high (>0.8) over extensive areas, reflecting stable hydrogeochemical conditions. Second, in southern and western China, low-fluoride zones (<0.2) displayed similarly strong consensus, where elevated topography and humid climates uniformly inhibited fluoride accumulation. Third, at basin margins and piedmont transitions, probabilities declined sharply over short lateral distances, producing narrow intermediate-probability fringes (0.4–0.6). Here, model uncertainty was not random but systematically structured by environmental complexity. RF concentrated its uncertainty zone (0.4–0.6) in the southern North China Plain (Fig. 5), where abrupt climate–soil transitions (evaporation decreasing from 1500 to 2000 mm to <1000 mm, precipitation increasing from 600 to 1000 mm to >1000 mm (Figure S2B, K); base saturation declining from > 80% to 50–80%, calcic alluvial soils transitioning to other soil types (Figure S2H, P)) created complex spatial configurations. LR produced broad, dispersed uncertainty bands across lowland margins and transitional boundaries—notably the Northeast Plain periphery, the Hetao–Ordos transition, and the Tarim Basin margins (Fig. 5)—reflecting its linear additive formulation's inability to represent threshold effects and nonlinear interactions among environmental gradients. ANN generated scattered uncertainty patches both at peripheries and within core areas (Northeast Plain), consistent with localized soil micro-heterogeneity (Figure S2P, Q) as a contributing factor to fluctuations in predicted probabilities. These divergent uncertainty patterns demonstrated that spatial variation in model confidence was itself spatially structured, clustering where environmental gradients steepened and multiple controls interacted (Fig. 5). This structured disagreement identifies zones of maximum information gain for independent ground-truthing, providing a rational basis for prioritizing future field sampling and validation efforts.

4.4. Population in predicted high-risk areas and public health prioritization

Quantifying the rural population within predicted geogenic high-risk zones provides a critical link between geogenic risk potential and public health prioritization. These estimates do not represent current exposure—extensive water-improvement projects have reduced fluoride levels in many historically endemic areas—but rather identify regions where natural hydrogeological conditions favor fluoride enrichment. Approximately 40.78–54.77 million rural residents reside in these zones, depending on the model applied (Figure S5), with the highest concentration in the North China Plain. The identification of provinces with persistent geogenic risk potential highlights the need for continued surveillance in densely populated groundwater-dependent regions. The predictive maps provide a prioritization framework for national drinking-water policy and surveillance in China. Consistent with the national Drinking Water Standard (GB 5749–2022), these risk layers can be integrated into existing monitoring systems to optimize sampling design and guide field investigation prioritization (Zhao et al., 2024). Incorporating these zones into surveillance plans enables rapid field validation, shifting practice from reactive monitoring toward proactive planning.

Management strategies should be stratified by risk category and intervention history. For established endemic regions with implemented water-improvement projects, continued surveillance is essential to verify long-term effectiveness. Where fluoride levels remain elevated despite intervention, source-switching—such as well siting in adjacent zones with lower predicted probability, or regional piped water transfer—should be prioritized over end-of-pipe treatment where feasible. Where source substitution is impractical, adsorption-based technologies or electrocoagulation may be applied (Carvalho and Oliveira, 2017; Karabulut et al., 2024). For model-identified candidate zones with no prior intervention history (e.g., Jianghan Plain, Sanjiang Plain, Junggar and Tarim basin rims), targeted field verification must precede any remediation decision, as these areas represent predicted geogenic potential rather than confirmed contamination. Decision-makers can stratify responses by overlaying geogenic risk potential with local constraints. However, any response should follow ground-truthing that validates predicted high-risk zones as areas with actual elevated fluoride levels.

4.5. Strengths, limitations, and future perspectives

The big-data analytical methodology established here—integrating extensive observational datasets, mechanistically informed predictors, and a three-model comparative framework—can be adapted to various environmental pollutants and health hazards. The use of existing national census data is standard practice for continental-scale geogenic risk mapping, as equivalent spatial coverage is not feasible for individual research teams (Cao et al., 2022; Podgorski and Berg, 2022). Nonetheless, several limitations should be considered when interpreting the findings. First, the analysis was based on cross-sectional spatial data and did not account for temporal variability in groundwater fluoride concentrations. Second, the lack of well-depth information constrained the ability to characterize vertical heterogeneity within aquifers. Third, population estimates were derived from the spatial overlap between predicted high-risk areas and rural population distributions, without accounting for variations in household water sources and usage patterns. Fourth, validation relied on spatial cross-validation and historical village records rather than independent post-hoc field sampling; targeted ground-truthing in model-predicted candidate zones would strengthen confidence in extrapolation to unmonitored regions. Future research should incorporate three-dimensional hydrogeological data to better represent vertical processes and include additional environmental and hydrogeochemical variables to further improve model performance. The integration of household-level water use data would also enhance the accuracy of population estimates. Moreover, coupling machine learning approaches with process-based hydrogeochemical models may provide a more comprehensive understanding of fluoride transport mechanisms and further strengthen predictive capability.

5. Conclusion

This study developed the first spatially cross-validated multi-model framework (ANN, RF, LR) for national-scale high-fluoride groundwater prediction in China, using 138,180 site records and 38 environmental predictors. The comparative approach reduced algorithm-specific bias while achieving strong predictive performance (PR_AUC > 0.95) and spatial generalizability. The results revealed that predicted geogenic risk is primarily concentrated in low-altitude plains and arid to semi-arid regions, where evaporation, soil properties, geological and volcanic conditions jointly promote fluoride enrichment. Model consensus confirmed known endemic zones, while divergences revealed new candidate hotspots (Jianghan Plain, Sanjiang Plain, Junggar and Tarim basin rims) awaiting field validation. An estimated 40.78–54.77 million rural residents reside in predicted high-risk areas, highlighting the North China Plain as the priority surveillance region. This scalable, spatially cross-validated tool supports targeted national groundwater safety management and offers a replicable approach for similar fluoride-affected regions.

Institutional review board statement

The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the National Institute for Endemic Disease Control and Prevention, Chinese Center for Disease Control and Prevention (protocol code hrbmuecdc20210303, date of approval is 1st March 2021).

CRediT authorship contribution statement

Yue Gao: Writing – review & editing, Supervision, Methodology. **Yanmei Yang:** Supervision, Project administration. **Peng Luo:** Resources. **Qingbo Wang:** Investigation. **Yunzhu Liu:** Validation. **Mengyao Su:** Writing – original draft, Visualization, Methodology. **Yanhui Gao:** Supervision, Methodology, Funding acquisition, Conceptualization. **Xin Zhang:** Software. **Chang Liu:** Formal analysis. **Xin Wang:**

Investigation. **Chao Zhang:** Data curation.

Informed consent statement

Informed consent was obtained from all subjects involved in the study.

Funding

This work was supported by the National Key R&D Program of China [2022YFC2503000].

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yanhui Gao reports financial support was provided by Key Research and Development Program of Heilongjiang. Yanhui Gao reports financial support was provided by National Natural Science Foundation of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We sincerely thank all study participants and research staff who participated in this work.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ecoenv.2026.120627.

Data availability

Data will be made available on request.

References

Abiye, T., Bybee, G., Leshomo, J., 2018. Fluoride concentrations in the arid Namaqualand and the Waterberg groundwater, South Africa: Understanding the controls of mobilization through hydrogeochemical and environmental isotopic approaches. Groundw. Sustain. Dev. 6, 112–120.

Alcaine, A.A., Schulz, C., Bundschuh, J., Jacks, G., Thunvik, R., Gustafsson, J.-P., et al., 2020. Hydrogeochemical controls on the mobility of arsenic, fluoride and other geogenic co-contaminants in the shallow aquifers of northeastern La Pampa Province in Argentina. Sci. Total. Environ. 715, 136671.

Ali, S., Mishra, P., Brindha, K., Gutiérrez, M., Kumar, R., Sunkari, E.D., et al., 2026. Global groundwater contamination by geogenic fluoride. Environ. Geochem. Health 48 (9), 410.

Barzegar, R., Asghari Moghaddam, A., Adamowski, J., Fijani, E., 2017. Comparison of machine learning models for predicting fluoride contamination in groundwater. Stoch. Environ. Res. Risk Assess. 31 (10), 2705–2718.

Cao, H., Xie, X., Wang, Y., Liu, H., 2022. Predicting geogenic groundwater fluoride contamination throughout China. J. Environ. Sci. 115, 140–148.

Carvalho, M.F., Oliveira, R.S., 2017. Natural production of fluorinated compounds and biotechnological prospects of the fluorinase enzyme. Crit. Rev. Biotechnol. 37 (7), 880–897.

Durrani, T.S., Farooqi, A., 2021. Groundwater fluoride concentrations in the watershed sedimentary basin of Quetta Valley, Pakistan. Environ. Monit. Assess. 193 (10), 644.

Gupta, P.K., Maiti, S., 2022. Enhancing data-driven modeling of fluoride concentration using new data mining algorithms. Environ. Earth Sci. 81.

He, X., Li, P., Ji, Y., Wang, Y., Su, Z., Elumalai, V., 2020. Groundwater Arsenic and Fluoride and Associated Arsenicosis and Fluorosis in China: Occurrence, Distribution and Management. Expo. Health 12 (3), 355–368.

Huang, L., Sun, Z., Zhou, A., Bi, J., Liu, Y., 2022. Source and enrichment mechanism of fluoride in groundwater of the Hotan Oasis within the Tarim Basin, Northwestern China. Environ. Pollut. 300, 118962.

Jia, Y., Xi, B., Jiang, Y., Guo, H., Yang, Y., Lian, X., et al., 2018. Distribution, formation and human-induced evolution of geogenic contaminated groundwater in China: A review. Sci. Total. Environ. 643, 967–993.

Karabulut, B.Y., Derin, P., Demir Yetiş, A., Yeşilnacar, M.İ., 2024. Health risk assessment in an area of dental fluorosis disease from high fluoride drinking water: a case study from southeastern Türkiye. Int. J. Environ. Health Res. 34 (5), 2299–2314.

Kazapoe, R.W., Yahans Amuah, E.E., Dankwa, P., Fynn, O.F., Addai, M.O., Berdie, B.S., et al., 2024. Fluoride in groundwater sources in Ghana: A multifaceted and country-wide review. Heliyon 10 (13), e33744.

Li, P., He, X., Li, Y., Xiang, G., 2019. Occurrence and health implication of fluoride in groundwater of loess aquifer in the Chinese loess plateau: a case study of Tongchuan, Northwest China. Expo. Health 11 (2), 95–107.

Li, Y., Bi, Y., Mi, W., Xie, S., Ji, L., 2021. Land-use change caused by anthropogenic activities increase fluoride and arsenic pollution in groundwater and human health risk. J. Hazard. Mater. 406, 124337.

Ling, Y., Podgorski, J., Sadiq, M., Rasheed, H., Eqani, S.A.M.A.S., Berg, M., 2022. Monitoring and prediction of high fluoride concentrations in groundwater in Pakistan. Sci. Total. Environ. 839, 156058.

Liu, J., Ma, Y., Gao, Z., Zhang, Y., Sun, Z., Sun, T., Fan, H., Wu, B., Li, M., Qian, L., 2022. Fluoride contamination, spatial variation, and health risk assessment of groundwater using GIS: a high-density survey sampling in Weifang City, North China. Environ. Sci. Pollut. Res. 29 (23), 34302–34313.

Lu, M.Y., Liu, Y., Liu, G.J., Li, Y.L., Xu, J.Z., Wang, G.Y., 2023. Spatial distribution characteristics and prediction of fluorine concentration in groundwater based on driving factors analysis. Sci. Total. Environ. 857, 159415.

Mohammadi, A.A., Ghaderpoori, M., Yousefi, M., Rahmatipoor, M., Javan, S., 2016. Prediction and modeling of fluoride concentrations in groundwater resources using an artificial neural network: a case study in Khaf. Environ. Health Eng. Manag. 3 (4), 217–224.

Nafouanti, M.B., Li, J., Mustapha, N.A., Uwamungu, P., AL-Alimi, D., 2021. Prediction on the fluoride contamination in groundwater at the Datong Basin, Northern China: Comparison of random forest, logistic regression and artificial neural network. Appl. Geochem. 132, 105054.

Nakaya, S., Takada, R., Yasumoto, J., Masuda, H., Yoshitani, J., et al., 2023. Effect of groundwater residence time on geogenic fluoride release into groundwater in the Mt. Meru slope area, Tanzania, the Great Rift Valley, East Africa. J. Contam. Hydrol. 253, 104125.

Nawrin, N., Arnaud, E., Priebe, E., 2024. Geospatial analysis of groundwater arsenic and fluoride in Quaternary aquifers of southern Ontario, Canada. Sci. Total. Environ. 923, 171485.

Nordstrom, D.K., 2022. Fluoride in thermal and non-thermal groundwater: Insights from geochemical modeling. Sci. Total. Environ. 824, 153606.

Podgorski, J., Berg, M., 2022. Global analysis and prediction of fluoride in groundwater. Nat. Commun. 13 (1), 4232.

Podgorski, J.E., Labhasetwar, P., Saha, D., Berg, M., 2018. Prediction Modeling and Mapping of Groundwater Fluoride Contamination throughout India. Environ. Sci. Technol. 52 (17), 9889–9898.

Rosecrans, C.Z., Belitz, K., Ransom, K.M., Stackelberg, P.E., McMahon, P.B., 2022. Predicting regional fluoride concentrations at public and domestic supply depths in basin-fill aquifers of the western United States using a random forest model. Sci. Total. Environ. 806, 150960.

Rubio, C., Rodríguez, I., Jaudenes, J.R., Gutiérrez, A.J., Paz, S., Burgos, A., et al., 2020. Fluoride levels in supply water from a volcanic area in the Macaronesia region. Environ. Sci. Pollut. Res. 27 (11), 11587–11595.

Sarkar, S., Mukherjee, A., Chakraborty, M., Quamar, M.T., Duttagupta, S., Bhattacharya, A., 2022. Prediction of elevated groundwater fluoride across India using multi-model approach: insights on the influence of geologic and environmental factors. Environ. Sci. Pollut. Res. 30 (11), 31998–32013.

Senarathne, S., Jayawardana, J.M.C.K., Chandrajith, R., 2022. Influence of climate on groundwater fluoride in different climatic domains in a hard rock terrain of Sri Lanka: implications to community health. Environ. Geochem. Health 44 (10), 3677–3686.

Solanki, Y.S., Agarwal, M., Gupta, A.B., Gupta, S., Shukla, P., 2022. Fluoride occurrences, health problems, detection, and remediation methods for drinking water: A comprehensive review. Sci. Total. Environ. 807, 150601.

Subbaiah, M.P., Yusuf, M., Ahn, Y., Park, H.B., Choi, J., Amin, M.A., Yadav, K.K., Jeon, B.-H., 2023. Fluoride occurrence in environment, regulations, and remediation methods for soil: A comprehensive review. Chemosphere 324, 138334.

Sun, Junliang, Chu, N., Zhao, B., Sun, Jianchong, 2026. National-scale classification and quantitative contribution of fluoride enrichment mechanisms in China's groundwater. Water Res. 297, 125664.

Virchow, L., Wilke, F., Yirgu, G., Neumann, T., Regenspurg, S., 2026. Fluoride distribution in volcanic rocks of the Aluto volcano complex, Main Ethiopian Rift. J. Afr. Earth Sci. 234, 105946.

Wang, C., Gao, Y., Wang, W., Zhao, L., Zhang, W., Han, H., et al., 2012. A national cross-sectional study on effects of fluoride-safe water supply on the prevalence of fluorosis in China. BMJ Open. 2 (5), e001564.

Wang, M., Li, X., He, W., Li, J., Zhu, Y., Liao, Y.-L., et al., 2019. Distribution, health risk assessment, and anthropogenic sources of fluoride in farmland soils in phosphate industrial area, southwest China. Environ. Pollut. 249, 423–433.

Wang, W., Li, Zhou, Su, H., Xiao, J., Han, F., Li, Zhi, 2022. Spatial and seasonal variability, control factors and health risk of fluoride in natural water in the Loess Plateau of China. J. Hazard. Mater. 434, 128897.

Xu, P., Bian, J., Li, Y., Wu, J., Sun, X., Wang, Y., 2022. Characteristics of fluoride migration and enrichment in groundwater under the influence of natural background and anthropogenic activities. Environ. Pollut. 314, 120208. https://doi.org/10.1016/j.envpol.2022.120208.

Yang, B., Wang, F., Yang, X., Cheng, J., Chen, J., Li, B., et al., 2025. The Systematic Analysis of the Impact of China Water Improvement and Fluoride Reduction Project on Children. Biol. Trace Elem. Res. 203 (10), 5018–5033.

Yang, Y., Zhang, R., Deji, Y., Li, Y., 2024. Hotspot mapping and risk prediction of fluoride in natural waters across the Tibetan Plateau. J. Hazard. Mater. 465, 133510.

Yuan, R., Li, Z., Guo, S., 2023. Health risks of shallow groundwater in the five basins of Shanxi, China: Geographical, geological and human activity roles. Environ. Pollut. 316, 120524.

Zhang, M., Wei, J., Wang, X., Ma, T., Li, F., Liu, J., et al., 2025. Spatial Distribution and Genesis of Fluoride in Groundwater, Qingshui River Plain, China. Water 17 (14), 2134.

Zhao, L., Li, Z., Li, M., Sun, H., Wei, W., Gao, L., et al., 2024. Spatial-Temporal Analysis of Drinking Water Type of Endemic Fluorosis — China, 2009–2022. China CDC Wkly 6 (2), 25–29.

Zhao, P., Zhang, S., Xu, K., Zhao, Y., Shen, P., Zhu, L., et al., 2023. Fluorine in shallow groundwater in China: A review of distribution, occurrence and environmental effects. Front. Earth Sci. 10, 1084890.

Zhu, J., Liu, R., Guo, H., Chen, J., Ning, D., Zang, X., 2025. Distribution Characteristics and Enrichment Mechanisms of Fluoride in Alluvial–Lacustrine Facies Clayey Sediments in the Land Subsidence Area of Cangzhou Plain, China. Water 17 (19), 2887.