ARTICLE

Investigating the association between groundwater contaminants and hypertension risk in India: a machine learning-based analysis

Sourav Biswas[1], Aparajita Chattopadhyay[1]✉, Kathrin Schilling[2] and Ayushi Das[3]

BACKGROUND: One-fourth of Indians are hypertensive, and the majority relies on groundwater for drinking. But the role of groundwater physicochemical properties and contamination in hypertension remains understudied.
OBJECTIVE: The study investigates the association between physicochemical groundwater characteristics andcontaminants and hypertension risk in India.
DATA: This study used data from the fifth round of the National Family Health Survey (NFHS-5 collected 2019–2021), including health, socio-demographics, and food and dietary information ($n = 712{,}666$ individuals). The physicochemical characteristics of groundwater data were derived from the Central Groundwater Board (CGWB, 2019–2021). This groundwater data from raster maps was linked to NFHS-5 records using cluster shapefiles and merging them with individual records via cluster IDs.
METHODS: Bivariate and multivariable regressions were used to identify factors associated with hypertension at the individual level. Moran's I statistics, Local Indicator of Spatial Association (LISA) cluster maps, and the Spatial Error Model (SEM) were used at district levels to investigate the spatial association. Machine learning models, including Artificial Neural Networks (ANN), Random Forest and Extreme Gradient Boosting (XGBoost), were used to predict hypertension risk zones.
RESULTS: Physicochemical drinking water composition is a key factor in hypertension risk. Elevated groundwater pH (>8.5, Adjusted Odds Ratio (AOR): 2.12), electrical conductivity (>300 µS/cm, AOR: 1.06), sulphate (>200 mg/L, AOR: 1.16), arsenic (>0.01 mg/L, AOR: 1.09), nitrate (>45 mg/L, AOR: 1.07), and magnesium (>30 mg/L, AOR: 1.03) are associated to higher odds of hypertension. The Random Forest model demonstrated the highest predictive performance, with a coefficient of determination (R^2) of 0.9970, mean absolute error (MAE) of 0.0012, and mean squared error (MSE) of 0.0077. It effectively identified high-risk zones in the northwestern (Delhi, Punjab, Haryana, and Rajasthan) and eastern (West Bengal and Bihar) regions of India.
IMPACT:

- This study highlights how important groundwater quality is in determining the incidence of hypertension, pointing to groundwater physicochemical properties and contaminants such as electrical conductivity, sulphate, arsenic, nitrate, and magnesium as essential factors. Our research is the first of its kind to comprehensively map hypertension risk zones using machine learning models and geospatial analysis. The findings highlight that water quality is a modifiable risk factor, reinforcing the need for improved drinking water supply systems, regular water quality testing, and targeted interventions in high-risk regions. This study emphasizes the importance of intersectoral collaborations to enhance public health outcomes.

Keywords: Hypertension; Non-communicable diseases; Groundwater; Drinking water; Contaminants; Environmental health

Journal of Exposure Science & Environmental Epidemiology; https://doi.org/10.1038/s41370-025-00776-0

INTRODUCTION

Non-communicable diseases (NCDs) account for 41 million deaths globally each year, representing 74% of all deaths [1]. Of these, 17 million occurr before age 70, predominantly in low- and middle-income countries (LMICs), which accounts for 86% of these premature deaths and 77% of all NCD-related deaths [1]. Hypertension, also known as high blood pressure, is the most common NCD, affecting around 1.13 billion people globally [2]. It is particularly prevalent in LMICs [2]. Hypertension is diagnosed when the systolic blood pressure is 140 mmHg or higher or the diastolic blood pressure is 90 mmHg or higher, written as '140/90 mmHg' [3]. In 2007, the Global Action Plan (GAP) was established to help prevent and handle NCDs like hypertension [4]. This plan also supports Sustainable Development Goal

[1]Department of Population & Development, International Institute for Population Sciences, Mumbai, Maharashtra 400088, India. [2]Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, 650 West 168th Street, New York, NY, USA. [3]Department of Survey Research & Data Analytics, International Institute for Population Sciences, Mumbai, Maharashtra 400088, India. ✉email: aparajita@iipsindia.ac.in

Received: 28 November 2024 Revised: 17 April 2025 Accepted: 17 April 2025
Published online: 04 June 2025

(SDG) 3.4, which targets a 25% reduction in NCD-related deaths by 2025 [5].

In India, hypertension rates vary widely across states and demographic. The National Nutrition Monitoring Bureau study (NNMB) from 2007/2008 reported hypertension prevalence of 27.1% among men and 26.4% among women in tribal areas, with Odisha (50–54.4%) and Kerala (36.7–45%) having the highest prevalence and Gujarat the lowest (7–11.5%) [6]. A 2014/2015 study in rural areas of Kerala and Andhra Pradesh observed that 47.7% of overweight individuals and 39.6% of those with higher waist circumference were hypertensive [7]. Another cross-sectional study reported high hypertension prevalence in West Bengal (29.5%) and Kerala (28.9%) but lower in Madhya Pradesh (16%) and Uttar Pradesh (19%) [8]. In 2015/2016, hypertension among people aged 15–49 was 11.3%, with men having a 4% higher prevalence than women, and urban areas (12.5%) showing slightly higher rates than rural areas (10.6%). The prevalence across states varied, ranging from 8.2% in Kerala to 20.3% in Sikkim [9]. The increase in hypertension has increased over the past decade and this increase is likely to continue due to socioeconomic shifts, environmental factors, and lifestyle changes [2].

The association of physicochemical groundwater composition, dietary habits, and hypertension risk is important but complex. Groundwater is the primary drinking water source for 85% of India's population [10], yet excessive usage raises sustainability and environmental concerns [11]. Dietary patterns have been associated with various other health outcomes, including type 2 diabetes [12–15], anaemia [16–19], and breast cancer [20–22]. Understanding the relationship between physicochemical groundwater composition and hypertension risk is crucial for effectively addressing the complexity of this health issue.

A seminal study in West Bengal found a strong association between groundwater arsenic levels and hypertension [23]. People in arsenic-endemic areas had nearly three times higher risk of hypertension compared to non-endemic areas [23]. A study in Bangladesh investigated the relationship between drinking water salinity and hypertension and found that people exposed to slight (slightly saltier than rainwater) and medium (moderately saltier than rainwater) or higher (much saltier than rainwater) levels of salinity have 1.28 times and 1.65 times greater risk of hypertension, respectively compared to those consuming drinking water with no salinity [24]. However, both studies primarily focused on single factors—arsenic contamination or salinity—without considering interactions with other physicochemical groundwater factors and dietary pattern that might increase hypertension risk.

Several studies have explored the socio-demographic characteristics, dietary patterns, and hypertension [9, 25]. Current research largely overlooked the combined effect of groundwater composition when used as primary drinking water source. Our study aims to clarify how physicochemical groundwater composition including known contaminants (e.g. arsenic) and dietary pattern together influence hypertension risk. We used machine learning techniques to predict hypertension risk zones across India based on identified environmental and health predictors. By identifying high-risk areas can help policymakers and healthcare professionals target interventions to mitigate the burden of hypertension and improve public health.

DATA AND METHODS

Data

Groundwater data. The analysis in this study is based on secondary data from the Central Groundwater Board (CGWB) under the Ministry of Jal Shakti, Department of Water Resources, River Development, and Ganga Rejuvenation. CGWB provides data from 29,065 monitoring sites across India collected between 2019 and 2021. The Indian states of Jammu & Kashmir, Ladakh, Sikkim, Nagaland, Manipur, Mizoram, Tripura and Lakshadweep have no data, and we excluded these states and union territories from our analysis. Table 1 provides an overview of the parametric values for drinking water compositions listed by the various countries and organizations [26–31].

Socio-demographic data. Food and dietary habits, socio-demographic, and hypertension data were leveraged from the National Family Health Survey (NFHS-5) conducted between 2019 and 2021 in India [32]. The survey included 28,43,917 individuals from 6,36,669 households across 28 states, 8 union territories, and 707 districts. Out of the total sample, 7,65,993 individuals aged 0–14 years were excluded because they were not tested, and 2,25,079 individuals declined blood sample testing. Additionally, 10,88,829 individuals who did not use groundwater for drinking and 39,010 individuals from hill states (where groundwater data was unavailable) were also excluded. The final sample size for analysis was 7,12,666 individuals (Fig. 1). Data included information on demographics, socioeconomic status, maternal and child health, reproductive health, and family planning. The analysis focused on men and women aged 15–95+ years, using data from the Person Recode (IAPR7DDT) accessed through the Demographic and Health Survey (DHS). To assess hypertension, blood pressure was measured using an Omron Blood Pressure Monitor. Three blood pressure readings were taken within a five-minute

Table 1. Acceptable limits for drinking water quality parameters from Bureau of Indian Standards (BIS), World Health Organization (WHO), European Union (EU), China, and United States Environmental Protection Agency (US. EPA).

Parameters	BIS (India)	WHO	EU	China	US EPA
Arsenic (mg/L)	0.01	0.01	0.01	0.01	0.01
Nitrate (mg/L)	45	50	50	20	10
Fluoride (mg/L)	1.0	1.5	1.5	1.0	4.0
Sulphate (mg/L)	200	500	250	250	250
Total Dissolved Solids (mg/L)	500	600	No limit specified	1000	500
Chloride (mg/L)	250	No guideline	250	250	250
Total Hardness (mg/L)	300	No guideline	No limit specified	450	No limit specified
Calcium (mg/L)	75	No guideline	No limit specified	No limit specified	No limit specified
Magnesium (mg/L)	30	No guideline	No limit specified	No limit specified	No limit specified
pH	6.5–8.5	6.5–8.5	6.5–9.5	6.5–8.5	6.5–8.5
Electrical Conductivity (μS/cm)	300	1500	2500	Not specified	No limit specified

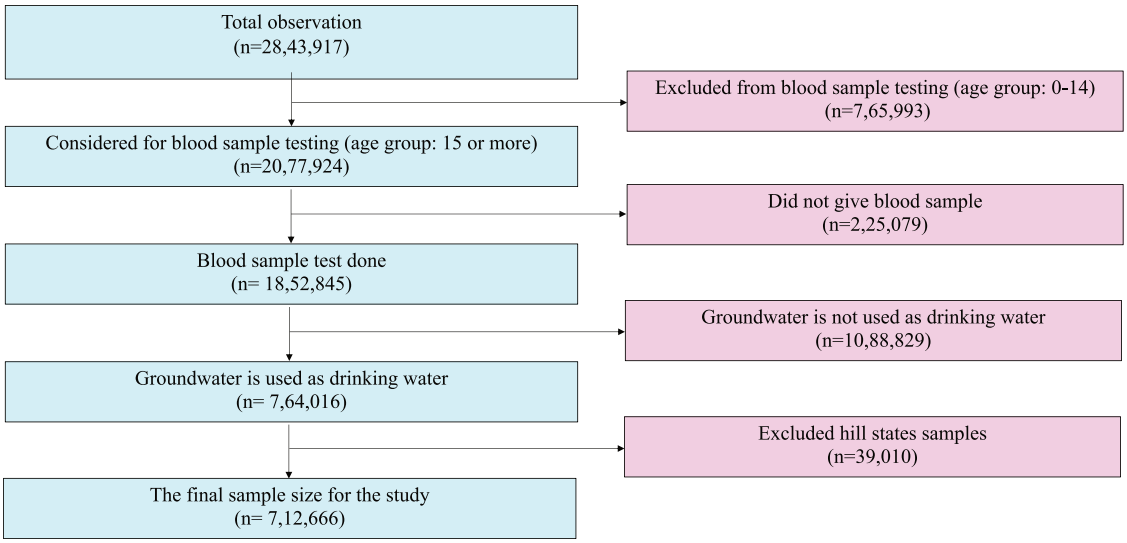

Fig. 1 Flow chart showing the steps to select representative samples.

interval. Individuals with an average systolic blood pressure (SBP) over 130 mm Hg and/or diastolic blood pressure (DBP) over 85 mm Hg were classified as hypertensive and advised to consult a doctor. The response rate for blood pressure measurement was 91.3% among women and 81.8% among men. Participants were asked if they had been diagnosed with hypertension on two or more occasions by a doctor, and if so, whether they were taking prescribed medication to lower their blood pressure.

Methods

Merging of CGWB water data and NFHS-5 data. The NFHS-5 survey collected data from 30,197 clusters, with socio-demographic and health information provided in Stata format and cluster locations available as a shapefile. The cluster identifier (v001) is common in both files. To integrate groundwater data with NFHS-5 data, we used groundwater quality data from 29,065 monitoring sites, which included latitude and longitude coordinates. First, the study applied the Inverse Distance Weighting (IDW) interpolation method [33] to create raster surfaces for the 11 groundwater quality parameters. Using ArcGIS, the study overlaid the NFHS-5 cluster shapefile onto these raster layers and extracted groundwater parameter values for each cluster. To account for spatial uncertainty, the spatial resolution of NFHS-5 clusters and the density of groundwater monitoring sites were considered, ensuring that the extracted values accurately represented local groundwater conditions. Finally, the extracted groundwater parameter values were merged with the NFHS-5 Stata file using the common cluster identifier (v001), creating the final dataset.

Outcome variable. For our study, we used hypertension as primary health outcome. Hypertension was diagnosed if systolic blood pressure (SBP) was ≥140 mmHg and/or diastolic blood pressure (DPB) was ≥90 mmHg, or if participants reported using medication to control blood pressure. The dichotomous variable, hypertension, was defined as 1 = hypertensive, otherwise = 0.

Explanatory variables. This study examined factors associated with hypertension. The factors included groundwater quality parameters (including contaminants), socioeconomic-demographic characteristics, and dietary habits. Groundwater quality parameters included pH, calcium, magnesium, sulphate, chloride, total hardness, total dissolved solids, electrical conductivity and contaminants nitrate, fluoride and arsenic. Socioeconomic-demographic variables included age, wealth index, residence, and geographic region. Dietary habits included consumption of milk, pulses/beans, vegetables, and non-vegetarian foods.

Statistical analysis. The study initially performed bivariate analysis techniques and Chi-square test to explore the relationships between hypertension and various factors, including groundwater quality parameters, socio-demographic characteristics, and dietary habits. Bivariate local Moran's I statistics, Local Indicators of Spatial Association (LISA) cluster map and scatter plots were generated using GeoDa 1.22 software to examine the association between hypertension and predictor variables at the district level. Moran's I, a metric of spatial autocorrelation akin to Pearson's correlation coefficient, was used to gauge spatial dependencies among districts. To delineate spatial connections, a spatial weight matrix was constructed using the queen contiguity method [34], which defines spatial relationships based on shared borders, ensuring a comprehensive representation of district-to-district connections. Each district was uniquely identified by a code, and the spatial weight matrix encapsulated the geographical interdependencies. The deliberate choice of using queen contiguity method was based on its ability to account for shared borders and provide a well-rounded spatial weight matrix for subsequent analyses. Significance in the spatial analysis was determined at the 0.05 significance level (p-values). This significance threshold was chosen to ensure the robust identification of spatial patterns and associations, facilitating meaningful conclusion about the spatial dynamics of hypertension risk and predictor variables across the diverse landscape of India. Moran's I calculations were performed utilizing the following formulas:

$$\text{Bivariate Local Moran's I}(I_{ij}) = \frac{(x_i - \bar{x})\left(\sum_{j=1}^{n} w_{ij}\,(y_i - \bar{y})\right)}{\sqrt{\sum_{j=1}^{n}(x_j - \bar{x})^2 \sum_{j=1}^{n}(y_j - \bar{y})^2}}$$

Here x_i: is the percentage of hypertension in district i, $\bar{x}$: is the mean of hypertension across all districts, y_i: is the percentage of predictor variable in district j, $\bar{y}$: is the mean of predictor variable across all districts, w_{ij}: is the spatial weight between districts i and j, and the sums are over all districts j in the dataset.

The Moran's I value ranges from −1, indicating perfect dispersion, to +1, indicating perfect correlation. A value of zero suggests a random spatial pattern. Negative values signify negative spatial autocorrelation, indicating dissimilarity among closely associated points, while positive values indicate positive spatial autocorrelation, signifying the clustering of points with similar attribute values in close proximity. The LISA cluster map serves as a visual

representation of spatial patterns, categorizing districts into four groups: "high-high," "low-low," "low-high," and "high-low," based on the prevalence of hypertension relative to neighbouring districts. "High-high" clusters depict districts with consistently high rates of hypertension prevalence, while "low-low" clusters denote districts with consistently low rates. Conversely, "low-high" clusters indicate districts with a low prevalence of hypertension surrounded by neighbours with high prevalence, and "high-low" clusters represent districts with a high prevalence of hypertension surrounded by neighbours with a low prevalence of hypertension. This classification identifies spatial clusters and geographical variations in hypertension prevalence across different regions.

Unadjusted and adjusted logistic regression models were used to ascertain groundwater's physicochemical characteristics and contaminants and other socio-demographic and dietary habit predictors with hypertension. Results were reported in terms of both unadjusted odds ratios (UORs) and adjusted odds ratios (AORs) with 95% confidence intervals (CIs). To evaluate the performance of the classification model, key diagnostic metrics were calculated, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), false positive rate (FPR), false negative rate (FNR), and overall accuracy. Sensitivity measured the proportion of true positive cases correctly identified, while specificity indicated the proportion of true negative cases correctly classified. PPV and NPV represented the probabilities that positive and negative predictions were correct, respectively. FPR captured the proportion of negative cases incorrectly classified as positive, and FNR reflected the proportion of positive cases missed by the model.

The study used regression models to examine the key variables associated with hypertension. Initially, spatial Ordinary Least Square (OLS) regression was used to assess the extent of autocorrelation in the error term. As the OLS regression confirmed spatial autocorrelation in its error term concerning the outcome variable, we further estimated spatial lag model (SLM) and Spatial Error Model (SEM). The underlying assumption of the spatial lag model posits that observations of the dependent variables are influenced by neighbouring areas, whereas the spatial error model incorporates the effects of variables not present in the regression model but impacting the outcome variable. The main difference between the two models is the use of spatial dependence in the error term, with the spatial lag model no considering spatial dependence [34]. Subsequently, based on the Akaike Information Criterion (AIC) values, the spatial error model seems to be the best-fitting model for our study. A typical spatial lag model can be written as follows:

$$Y_i = \delta \sum_{j \neq 1} W_{ij} Y_j + \beta X_j + \varepsilon_j$$

Here Y_i: denotes the prevalence of hypertension for the i^{th} district, δ: is the spatial autoregressive coefficient, W_{ij}: denotes the spatial weight of proximity between district i and j, Y_j: is the prevalence of hypertension in the j^{th} district, β: denotes the coefficient, X_j: is the predictor variable and ε_j: is the residual.

The Spatial Error Model takes into consideration the impact of omitted variables that are not directly incorporated into the model but could substantially affect the analysis. Formally, a Spatial Error Model (SEM) can be articulated as follows:

$$Y_i = \beta X_j + \lambda \sum_{j \neq 1} W_{ij} Y_j \varepsilon_j + \varepsilon_j$$

Here Y_i: denotes the prevalence of hypertension for the i^{th} district, λ: is the spatial autoregressive coefficient, W_{ij}: denotes the spatial weight of proximity between district i and j, Y_j: is the prevalence of hypertension in the j^{th} district, β: denotes the coefficient, X_j: is the predictor variable and ε_j: is the residual [35].

To examine spatial heterogeneity in the association between groundwater physicochemical parameters and contaminants and hypertension risk, the study employed Geographically Weighted Regression (GWR) model [36, 37]. Unlike traditional global regression models, GWR allows regression coefficients to vary across geographic locations, capturing localized relationships between independent variables and hypertension prevalence. The model is specified as:

$$Yi = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) X_{ki} + \varepsilon_i$$

Where Yi represents the prevalence of hypertension in districts i, X_{ki} includes groundwater physicochemical properties and contaminants (e.g., pH, Nitrate, Fluoride, Arsenic), $\beta_k(u_i, v_i)$ are location-specific coefficients, $\beta_o(u_i, v_i)$ is the intercept indicating baseline hypertension prevalence, and ε_i is the error term. The GWR model provides key outputs: regression coefficients (β_k), which indicate how the relationship between groundwater contamination and hypertension varies spatially; local R^2 values, measuring how well predictors explain hypertension prevalence at different locations [36]; standardized residuals, identifying areas where model predictions deviate from observed values; and the intercept (β_0), representing the baseline hypertension prevalence across regions. A higher local R^2 suggests stronger spatial associations, whereas lower values indicate weaker relationships, possibly due to unmeasured local factors. The model was implemented using an adaptive bandwidth selection based on the AIC for optimal performance [37].

The study utilized three supervised machine learning models—Artificial Neural Networks (ANN), Random Forest (RF), and Extreme Gradient Boosting (XGBoost)—to predict hypertension risk zones. ANN, inspired by the human brain's neural structure, uses interconnected nodes to identify patterns and generate predictions. ANN applied backpropagation and a sigmoid activation function to adjust weights iteratively until convergence in this study [38]. Random Forest combines multiple decision trees to reduce overfitting and improve accuracy, with trees operating independently in parallel [39]. XGBoost, an ensemble algorithm, sequentially builds decision trees to minimize prediction errors by optimizing a loss function through gradient descent, incorporating L1 and L2 regularization to control overfitting [40]. The dataset was divided into train and test set by 80:20 ratio. The machine learning models were fitted only on the train set and test set was kept for testing the models for accuracy. A 10-fold cross-validation was used to optimize the hyperparameters. The Artificial Neural Network (ANN) model was compiled using the Adam optimizer, with mean squared error (MSE) as the loss function and mean absolute error (MAE) as the performance metric. For the XGBoost (XGB) model, the XGBRegressor was initialized with the objective function reg:squarederror and a fixed random state [40] to ensure reproducibility. The Random Forest (RF) model was implemented using its default parameters, including 100 trees (n_estimators = 100) and squared error as the criterion (criterion = 'squared_error') for measuring the quality of splits. Using data from NFHS-5 and CGWB, the study predict hypertension in 28,501 cluster points in the Python package, creating the final map using QGIS 3.18.1. Figure 2 illustrates the conceptual framework for our machine-learning-based hypertension prediction model. The whole statistical analysis was conducted using StataSE 16 software. Figure 3 shows the conceptual framework that directed the analysis of our study.

RESULTS

Spatial distribution of physicochemical groundwater parameters and groundwater contaminants

The study analysed physicochemical groundwater parameters and groundwater contaminants from 29,065 CGBW monitoring wells,

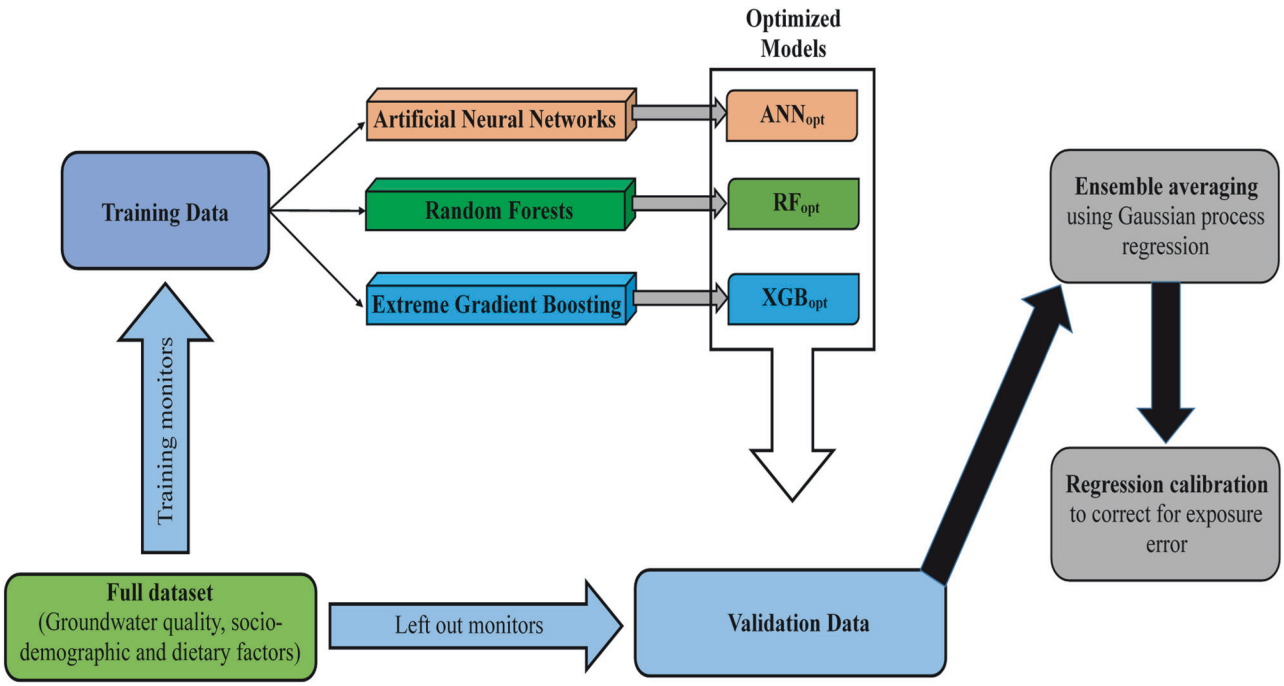

Fig. 2 Model development framework: machine-learning-based framework for developing the hypertension risk zone prediction model in India.

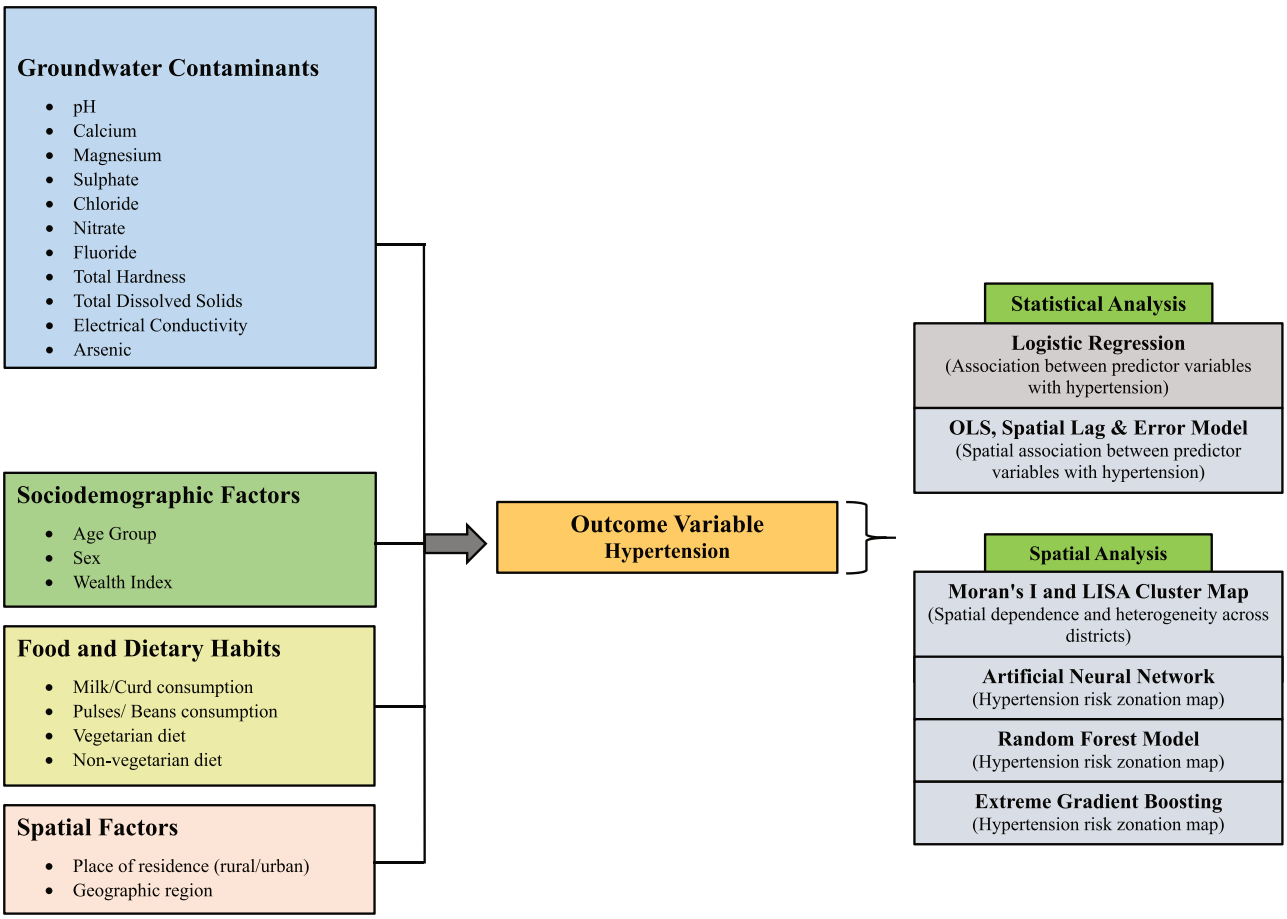

Fig. 3 Conceptual framework of the study.

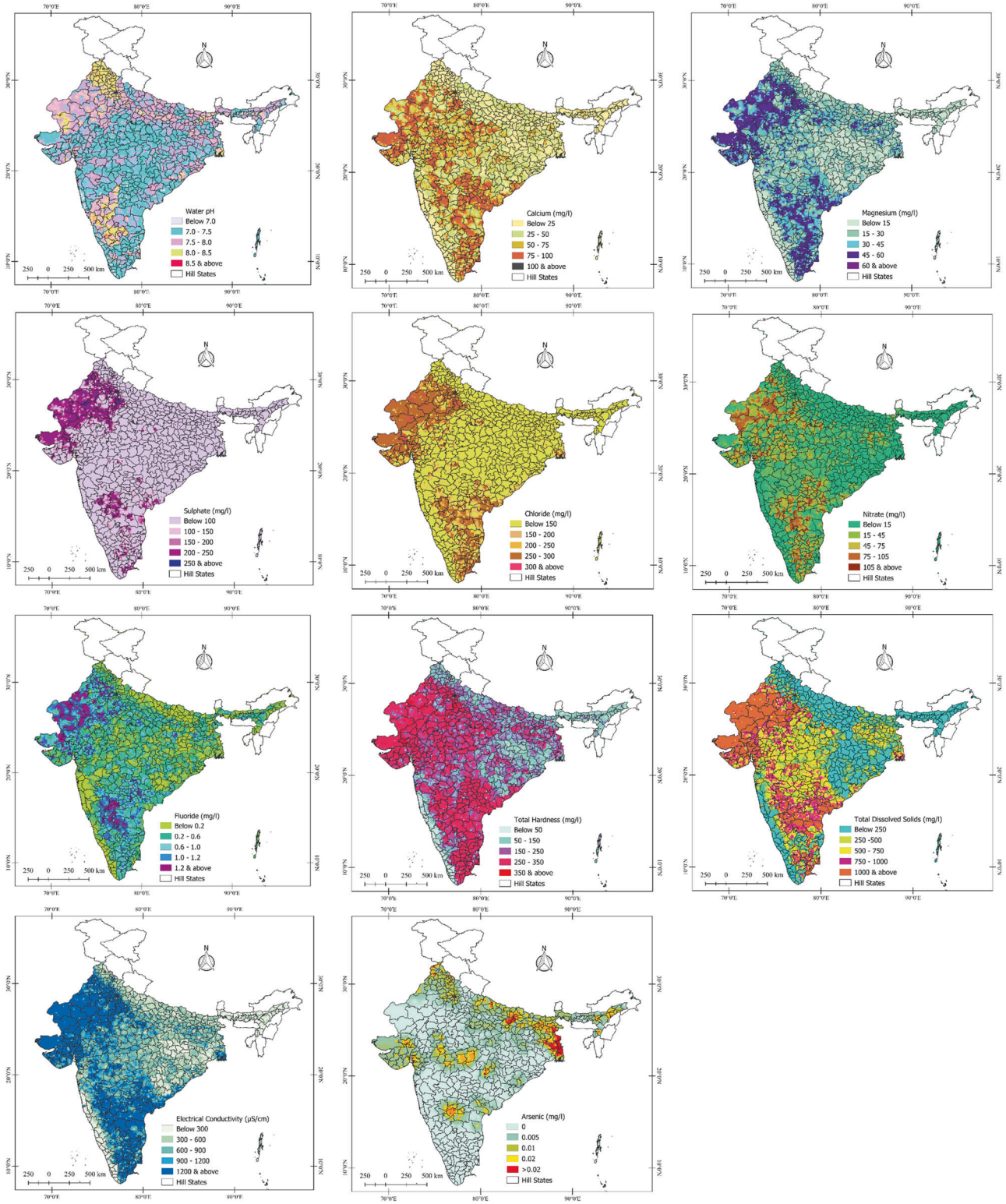

Fig. 4 Spatial distribution of various groundwater physicochemical properties and contaminants across India.

revealing distinct spatial distribution patterns (Fig. 4). Using IDW interpolation, the study identified spatial hotspots of various physicochemical groundwater parameters and groundwater contaminants, highlighting groundwater quality variations nationwide. High arsenic levels were observed in the northwestern, eastern, and northeastern regions, predominantly within the Ganga, Indus, and Brahmaputra river basins. High levels of nitrate, magnesium, calcium, fluoride, sulphate, chloride, total hardness, total dissolved solids, and electrical conductivity were detected in the northwest and southeastern parts of India, often exceeding Bureau of Indian Standards (BIS) thresholds (Table 1). The pH map indicated basic conditions in the northwestern, western, southwestern, and eastern regions, while slightly acidic conditions were found in central Indian regions, reflecting potential geochemical variations across these areas.

Table 2. Bivariate association between physicochemical groundwater parameters, groundwater contaminants and other socio-demographic factors with hypertension risk in India.

Selected background characteristics	Hypertension		
	Yes (1,52,368)	No (5,60,298)	*p*-value
Arsenic			
≤0.01 mg/L	20.96	79.04	0.0001
>0.01 mg/L	23.35	76.65	
Nitrate			
≤45 mg/L	21.14	78.86	0.0001
>45 mg/L	22.56	77.44	
Magnesium			
≤30 mg/L	21.32	78.68	0.0004
>30 mg/L	21.82	78.18	
Calcium			
≤75 mg/L	22.13	77.87	0.0001
>75 mg/L	20.26	79.74	
Fluoride			
≤1 mg/L	21.31	78.69	0.0006
>1 mg/L	21.45	78.55	
Sulphate			
≤200 mg/L	20.61	79.39	0.0002
>200 mg/L	22.87	77.13	
Chloride			
≤250 mg/L	21.24	78.76	0.0002
>250 mg/L	21.78	78.22	
Total hardness			
≤300 mg/L	21.21	78.79	0.0001
>300 mg/L	21.72	78.28	
Total dissolved solids			
≤500 mg/L	21.48	78.52	0.0003
>500 mg/L	21.28	78.72	
Electrical conductivity			
≤300 µS/cm	20.16	79.84	0.0004
>300 µS/cm	22.12	77.88	
Water pH			
≤6.5	22.55	77.45	0.0001
6.5–8.5	21.16	78.84	
>8.5	26.12	73.88	
Age groups			
15–24 years	4.79	95.21	0.0002
25–34 years	10.53	89.47	
35–44 years	19.77	80.23	
45–54 years	30.53	69.47	
55–64 years	39.96	60.04	
65 and above	48.79	51.21	
Sex			
Male	22.66	77.34	0.0001
Female	20.28	79.72	
Wealth Index			
Poorest	17.94	82.06	0.0001
Poorer	18.75	81.25	
Middle	20.93	79.07	
Richer	23.16	76.84	
Richest	26.07	73.93	
Milk/curd consumption			
Never	21.08	78.92	0.0004
Daily	23.12	76.88	
Weekly	20.84	79.16	
Occasionally	20.92	79.08	
Pulses/beans consumption			
Never	20.84	79.16	0.0002
Daily	22.36	77.64	
Weekly	21.02	78.98	
Occasionally	20.73	79.27	
Vegetarian diet			
Never	21.92	78.08	0.0002
Daily	21.13	78.87	
Weekly	21.05	78.95	
Occasionally	20.94	79.06	
Non-vegetarian diet			
Never	21.24	78.76	0.0001
Daily	22.32	77.68	
Weekly	21.32	78.68	
Occasionally	20.88	79.12	
Place of residence			
Urban	24.18	75.82	0.0001
Rural	20.45	79.55	
Geographic region			
North	22.77	77.23	0.0001
Central	20.65	79.35	
East	18.71	81.29	
Northeast	18.02	81.98	
West	20.01	79.99	
South	26.16	73.84	
Total	21.38	78.62	

Hypertension by background characteristics

Table 2 shows the bivariate analysis of physicochemical characteristics and contaminants of groundwater, socio-demographic factors, and hypertension. Hypertension affected 23.4% of individuals with arsenic levels above 0.01 mg/L, compared to 20.96% in areas at or below this level. Similarly, 22.56% of individuals were hypertensive in areas with nitrate levels above 45 mg/L, compared to 21.14% in areas below 45 mg/L nitrate. For calcium, hypertension prevalence was 20.26% in areas with levels above 75 mg/L, compared to 22.13% in areas at or below this threshold. Areas with total hardness exceeding 300 mg/L had 21.72% of individuals with hypertension, compared to 21.21% in areas with hardness at or below 300 mg/L. For water pH, 26.12% of individuals were hypertensive in areas with pH above 8.5, 21.16% in areas between 6.5 and 8.5, and 22.55% in areas at or below pH 6.5.

Table 3 presents the Bivariate Moran's I statistics, summarizing the spatial associations between hypertension prevalence and groundwater physicochemical properties and contaminants. Zonal statistics were used to compute the mean values of groundwater

Table 3. Moran's I statistics show the spatial association for hypertension and its correlates in India.

Indicators	Moran's I value	*p*-value
Arsenic	0.03	0.002
Nitrate	0.09	0.001
Magnesium	−0.03	0.024
Calcium	−0.01	0.031
Fluoride	0.03	0.003
Sulphate	0.05	0.002
Chloride	0.02	0.012
Total hardness	−0.01	0.046
Total dissolved solids	−0.13	0.001
Electrical conductivity	−0.02	0.032
Water pH	0.05	0.002

physicochemical properties and contaminants indicators at the district level, while hypertension data were derived from the NFHS-5 dataset. The updated results indicate positive spatial autocorrelations between hypertension and arsenic (0.03, Moran's I), nitrate (0.09), fluoride (0.03), sulphate (0.05), chloride (0.02), and water pH (0.05). In contrast, negative spatial autocorrelations were observed for magnesium (−0.03), calcium (−0.01), total hardness (−0.01), total dissolved solids (−0.13), and electrical conductivity (−0.02).

The LISA cluster analysis (Fig. 5) reveals significant high-high clusters, indicating regions where groundwater contaminants and hypertension prevalence are simultaneously elevated. Arsenic and hypertension high-high clusters are observed in 29 districts primarily located in Punjab, Haryana, and parts of Andhra Pradesh. Similarly, nitrate and hypertension high-high clusters are found in 54 districts, covering parts of Punjab, Haryana, Andhra Pradesh, Telangana, Tamil Nadu, and Karnataka. Magnesium and hypertension high-high clusters are identified in 44 districts, mainly in parts of Punjab, Telangana, Andhra Pradesh, Tamil Nadu, and Karnataka, while calcium and hypertension high-high clusters are present in 43 districts, covering similar regions. Fluoride and hypertension high-high clusters are noted in 57 districts, spanning Punjab, Andhra Pradesh, Telangana, Tamil Nadu, and Karnataka. Additionally, sulphate (44 districts), chloride (40 districts), total hardness (45 districts), total dissolved solids (38 districts), and electrical conductivity (47 districts) exhibit high-high clustering with hypertension, predominantly in Punjab, Andhra Pradesh, Telangana, Tamil Nadu, and Karnataka. pH and hypertension high-high clusters are observed in 38 districts, mainly in Punjab, Haryana, and Karnataka. These spatial patterns suggest a strong association between groundwater contamination and hypertension prevalence.

Factors associated with hypertension

Table 4 shows the odds ratios from the logistic regression model for factors associated with hypertension. Three logistic models are analysed: Model 1 includes only groundwater parameters as independent variables; Model 2 incorporates socio-demographic, food and dietary habits, and spatial factors; and Model 3, the adjusted model, encompasses all factors. The odds ratios highlight significant associations between groundwater quality parameters, contaminants, and hypertension. The groundwater parameters with the strongest effects on hypertension include water pH and sulphate levels. Individuals in areas with a water pH above 8.5 have 2.12 times higher odds of hypertension compared to those in areas with pH ≤6.5, making it the most impactful factor. Sulphate levels above 200 mg/L are associated with 1.16 times higher odds of hypertension. Arsenic levels above 0.01 mg/L increase hypertension odds by 1.09 times, and nitrate levels above 45 mg/L are linked to 1.07 times higher odds. Elevated electrical conductivity above 300 µS/cm increases hypertension odds by 1.06 times, while magnesium levels above 30 mg/L raise the odds by 1.03 times. In contrast, some groundwater parameters appear to reduce hypertension odds. Calcium levels above 75 mg/L are associated with 0.92 times lower odds, and total dissolved solids above 500 mg/L are linked to 0.94 times lower odds. Individuals in areas with a pH between 6.5 and 8.5 have 0.88 times lower odds of hypertension compared to areas with pH ≤6.5. The findings highlight the differential impacts of groundwater quality parameters on hypertension risk. Specifically, elevated pH and sulfate levels are found to have a significant influence, while higher concentrations of calcium and total dissolved solids seem to offer a protective effect.

Table 5 provides key performance metrics for a classification model, and the accompanying Receiver Operating Characteristic (ROC) curve (Fig. 6) helps visualize its overall predictive ability. The model's sensitivity (78.2%) indicates that it correctly identifies 78.2% of true positive cases, while its specificity (90.5%) reflects a strong ability to correctly classify negative cases. The positive predictive value (67.8%) shows that 67.8% of predicted positives are actual positives, whereas the negative predictive value (92.3%) reflects that 92.3% of predicted negatives are true negatives. The false positive (9.5%) and false negative rates (21.8%) indicate that the model has a relatively lower proportion of incorrect positive classifications but misses about 21.8% of actual positive cases. The overall accuracy (88.7%) suggests strong classification performance. The ROC curve further supports this by illustrating the trade-off between sensitivity and specificity at different thresholds. The area under the ROC curve (AUC = 0.7526) suggests that the model has moderate discriminative ability, meaning it performs significantly better than random classification (AUC = 0.5) but leaves room for improvement.

Spatial analysis of hypertension correlates in India: findings from spatial error model

Table 6 presents the spatial determinants and influential factors affecting hypertension in India. Following the establishment of significant bivariate spatial associations between the dependent and independent variables, an OLS model was fitted. The OLS model indicated that the residuals were spatially auto correlated. Subsequently, Spatial Lag and Error Models were applied to the data. Among the two estimated spatial models, the SEM exhibited the lowest AIC value, thus considered the appropriate model for the study. The Spatial Error Model confirmed that environmental groundwater factors such as arsenic, sulphate, and pH remained statistically significant predictors of hypertension risk. The SEM model provided the final spatial estimates of correlates of hypertension (Table 6). The Lambda value was 0.698, which is highly significant, indicating positive spatial autocorrelation of regions with a high prevalence of hypertension in India. The coefficient of arsenic was the highest (β = 2.296), followed by pH (β = 0.927) and sulphate (β = 0.107). The coefficient estimate for arsenic confirmed that a 10-point increase in the proportion of groundwater arsenic was associated with a 23-point increase in hypertension. Similarly, a 10-point increase in groundwater pH was associated with a 9.27-point increase in hypertension risk. Similarly, a 10-point increase in sulphate were associated with 1.07 points increase in hypertension prevalence.

Spatial variability in the association between groundwater physicochemical properties and contaminants with hypertension: geographically weighted regression results

The GWR model (Fig. 7) achieved an adjusted R^2 of 0.57, indicating that physicochemical groundwater parameters and contaminants explain 57% of the spatial variability in hypertension prevalence across Indian districts, with an AICc of 1237.48, suggesting a good model fit. The standardized residuals analysis shows that 505 out

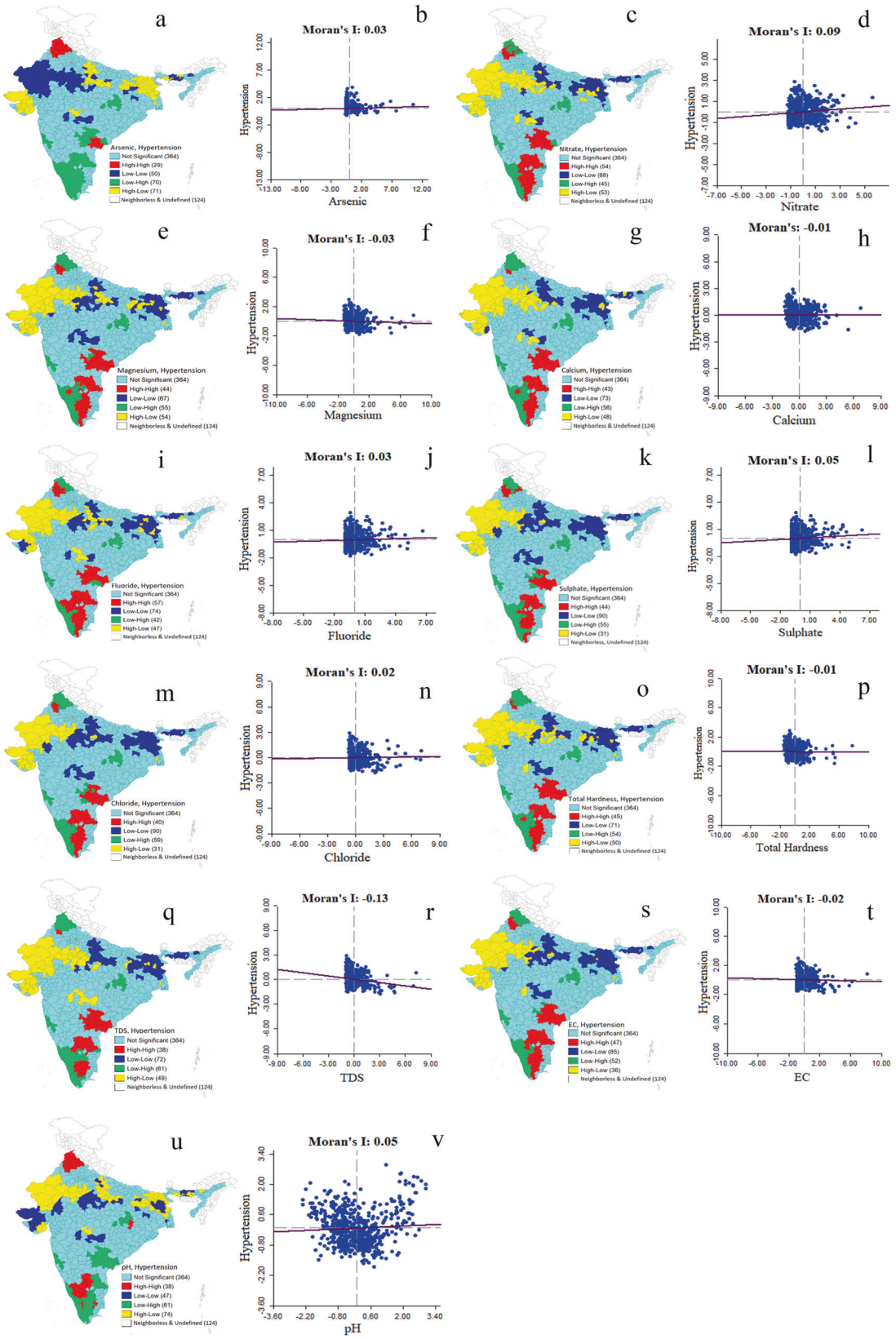

Fig. 5 Bivariate LISA cluster maps and scatter plots showing the spatial clustering of groundwater physicochemical properties and contaminants with hypertension in India. Bivariate LISA cluster maps and scatter plots showing the geographic clustering of: **a**, **b** arsenic, **c**, **d** nitrate, **e**, **f** magnesium, **g**, **h** calcium, **i**, **j** fluoride, **k**, **l** sulphate, **m**, **n** chloride, **o**, **p** total hardness, **q**, **r** total dissolved solids, **s**, **t** electrical conductivity, and **u**, **v** pH with hypertension at the district level in India. Note: In the legend, the numbers shown in parentheses indicate the number of districts falling into each particular group.

Table 4. Unadjusted and adjusted effects (odds-ratio, OR) estimate the association between groundwater physicochemical parameters and groundwater contaminants, and other socio-demographic factors with hypertension in India.

Selected background characteristics **Observations = 7,12,666**	**Model 1** **OR [95% CI]**	**Model 2** **OR [95% CI]**	**Model 3** **OR [95% CI]**
Arsenic			
≤0.01 mg/L	ref		ref
>0.01 mg/L	1.12***[1.04,1.16]		1.09***[1.05,1.15]
Nitrate			
≤45 mg/L	ref		ref
>45 mg/L	1.05**[1.01,1.09]		1.07**[1.03,1.12]
Magnesium			
≤30 mg/L	ref		ref
>30 mg/L	1.06**[1.01,1.11]		1.03**[1.00,1.08]
Calcium			
≤75 mg/L	ref		ref
>75 mg/L	0.89**[0.83,0.96]		0.92**[0.85,0.99]
Fluoride			
≤1 mg/L	ref		ref
>1 mg/L	0.98[0.96,1.04]		1.01[0.97,1.06]
Sulphate			
≤200 mg/L	ref		ref
>200 mg/L	1.33***[1.02,1.43]		1.16**[1.04,1.28]
Chloride			
≤250 mg/L	ref		ref
>250 mg/L	1.08*[1.01,1.16]		1.03[0.91,1.11]
Total Hardness			
≤300 mg/L	ref		ref
>300 mg/L	1.02**[1.00,1.10]		1.04[0.97,1.11]
Total Dissolved Solids			
≤500 mg/L	ref		ref
>500 mg/L	0.87**[0.80,0.94]		0.94**[0.85,0.96]
Electrical Conductivity			
≤300 μS/cm	ref		ref
>300 μS/cm	1.02*[1.00,1.10]		1.06*[1.01,1.09]
Water pH			
≤6.5	ref		ref
6.5 - 8.5	0.91**[0.73,0.92]		0.88**[0.74,0.97]
>8.5	2.38**[2.02,2.46]		2.12**[1.97,2.22]
Age group			
15-24 years		ref	ref
25-34 years		2.30***[2.25,2.34]	2.29***[2.25,2.34]
35-44 years		4.91***[4.83,5.00]	4.91***[4.82,4.99]
45-54 years		8.47***[8.32,8.62]	8.47***[8.32,8.62]
55-64 years		12.84***[12.61,13.08]	12.86***[12.62,13.10]
65 and above		18.83***[18.48,19.19]	18.81***[18.46,19.17]
Sex			
Male		ref	ref
Female		0.85***[0.85,0.86]	0.85***[0.85,0.86]
Wealth Index			
Poorest		ref	ref
Poorer		1.09***[1.07,1.10]	1.09***[1.08,1.11]
Middle		1.24***[1.22,1.26]	1.24***[1.22,1.26]
Richer		1.40***[1.38,1.42]	1.40***[1.37,1.42]
Richest		1.56***[1.53,1.58]	1.53***[1.50,1.56]

Table 4. continued

Selected background characteristics Observations = 7,12,666	Model 1 OR [95% CI]	Model 2 OR [95% CI]	Model 3 OR [95% CI]
Milk/Curd consumption			
Never		ref	ref
Daily		0.86***[0.85,0.88]	0.88***[0.86,0.89]
Weekly		0.89***[0.87,0.91]	0.90***[0.88,0.92]
Occasionally		0.96***[0.94,0.98]	0.96***[0.94,0.98]
Pulses/beans consumption			
Never		ref	ref
Daily		1.02[0.96,1.10]	1.02[0.95,1.10]
Weekly		1.00[0.94,1.08]	1.01[0.94,1.08]
Occasionally		1.01[0.94,1.08]	1.01[0.94,1.09]
Vegetarian diet			
Never		ref	ref
Daily		0.97[0.90,1.06]	0.97[0.89,1.05]
Weekly		0.95[0.88,1.03]	0.95[0.88,1.03]
Occasionally		0.91*[0.84,0.99]	0.91*[0.84,0.99]
Non-vegetarian diet			
Never		ref	ref
Daily		1.14***[1.08,1.21]	1.12***[1.06,1.19]
Weekly		1.13***[1.11,1.14]	1.11***[1.10,1.13]
Occasionally		1.10***[1.08,1.11]	1.08***[1.07,1.09]
Place of residence			
Urban		ref	ref
Rural		0.96***[0.95,0.97]	0.95***[0.93,0.96]
Geographic region			
North		ref	ref
Central		1.00[0.98,1.01]	0.95***[0.93,0.97]
East		0.81***[0.79,0.82]	0.79***[0.77,0.80]
Northeast		0.87***[0.85,0.90]	0.85***[0.82,0.87]
West		0.86***[0.84,0.87]	0.86***[0.84,0.87]
South		1.08***[1.06,1.09]	1.08***[1.06,1.10]

Model 1: Unadjusted estimates using only groundwater contaminants and physicochemical properties.
Model 2: Unadjusted estimates using only age, sex, wealth index, dietary habits, place of residence, and geographic region.
Model 3: Fully adjusted, including groundwater contaminants, physicochemical properties, age, sex, wealth index, dietary habits, place of residence, and geographic region.
CI Confidence Interval, *OR* Odds-Ratio, *ref* Reference Category.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Table 5. Sensitivity, specificity, and classification performance metrics.

Metric	Value (%)
Sensitivity (true positive rate)	78.2
Specificity (true negative rate)	90.5
Positive predictive values	67.8
Negative predictive values	92.3
False positive rate	9.5
False negative rate	21.8
Correctly classified	88.7

of 587 districts (86%) fall within the range of −1.5 to +1.5, indicating that the model performs well in most regions, though some districts exhibit higher deviations. The local R^2 values reveal distinct spatial variations, with higher values (0.55–0.65) concentrated in north-western regions such as Punjab, Haryana, and Delhi. Indicating a strong association between groundwater parameters and hypertension. Lower values (0.15–0.25) are observed in central India, including Madhya Pradesh and Chhattisgarh, suggesting weaker correlations and the possible influence of other environmental or socio-economic factors. The intercept values further highlight spatial heterogeneity, with higher intercepts (>0.50) in northeastern states like Assam and Tripura and parts of West Bengal and Odisha, indicating a higher baseline prevalence of hypertension even after accounting for groundwater parameters. In contrast, lower intercept values (<0.20) are observed in southern and western regions, including Kerala, Tamil Nadu, and Gujarat, suggesting a lower baseline prevalence.

Figure 8 illustrates the spatial distribution of GWR coefficients, revealing distinct regional patterns for physicochemical groundwater parameters and contaminants. For arsenic, the highest coefficients (>0.30) are observed in the western region (Rajasthan and parts of Gujarat), while lower coefficients (−0.20 to −0.30) appear in southern states like Kerala and Tamil Nadu. For nitrate,

high coefficients (0.20–0.30) are seen in central and southern regions, while northern areas, especially the Indo-Gangetic plains, show lower coefficients (−0.10 to −0.20). In the case of fluoride, high coefficients (>0.20) are found in eastern parts, including Bihar, Jharkhand, Odisha, and West Bengal, while western and central regions also show moderately high values. For sulphate, coefficients greater than 0.20 are seen across northern states (Uttar Pradesh), eastern states (Bihar, Jharkhand, Odisha, West Bengal), northeastern states (Assam, Tripura), and western states (Rajasthan, Punjab, Haryana, Gujarat). For groundwater total dissolved solids, most regions show negative coefficients (<−0.20), indicating widespread negative spatial association in most parts of India. For chloride, most Indian regions show positive coefficients, except for southern Kerala, parts of Tamil Nadu, and northeastern regions such as Assam and Tripura, where negative coefficients dominate. Total hardness shows positive coefficients in Gujarat, Maharashtra, Karnataka, Goa, Kerala, parts of Tamil Nadu and Andhra Pradesh, and western Uttar Pradesh. For calcium, positive coefficients are primarily observed across the northern states (Punjab, Haryana, Uttar Pradesh) and eastern states (Bihar, Jharkhand, and West Bengal). Magnesium also displays positive coefficients in Gujarat, Maharashtra, Karnataka, Goa, Kerala, parts of Tamil Nadu, and Andhra Pradesh, indicating spatial clustering in western and southern India. For pH, positive coefficients are found across the northern states (Punjab, Haryana, Himachal Pradesh, and Uttar Pradesh), northwestern states (Rajasthan and Gujarat), and northeastern states (Assam, Meghalaya, and Arunachal Pradesh). Finally, electrical conductivity shows positive coefficients in northern states (Punjab, Haryana, Uttar Pradesh, Bihar) and northeastern states (Assam, Arunachal Pradesh), highlighting distinct spatial clustering in these regions.

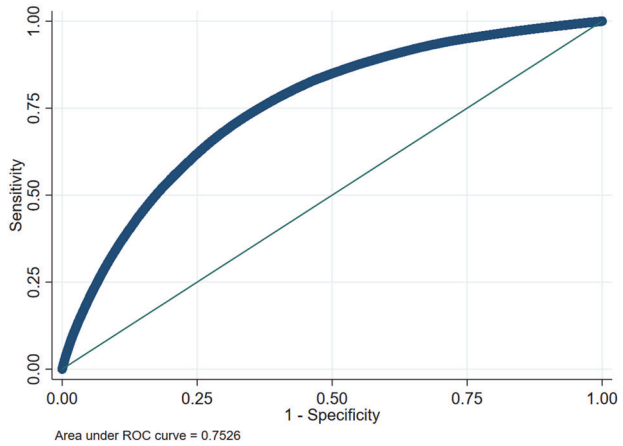

Fig. 6 Receiver operating characteristic (ROC) curve showing sensitivity and specificity for hypertension classification.

Hypertension risk zonation maps using machine learning techniques

The three machine learning models were evaluated by mean squared error (MSE), mean absolute error (MAE), and the R^2 score (Table 7). The RF (MAE = 0.0012, MSE = 0.0077, R^2 = 0.9970) regressor was the best-fit model, with ANN (MAE = 0.0236, MSE = 0.0437, R^2 = 0.9038) being the second best and XGBoost (MAE = 0.0977, MSE = 0.1342, R^2 = 0.0930) was the least performing model. Figure 9 illustrates the hypertension risk zones across India. The hypertension risk ranges from 0 to 1, where 0 signifies no risk, and 1 indicates a very high risk of hypertension. Areas of high to very high risk of hypertension are predominantly in the north-western regions of India including Delhi, Punjab, Haryana, and Rajasthan. The western and southern parts of India, including Gujarat, Telangana, Andhra Pradesh, Karnataka, and Tamil Nadu, exhibit a medium to high risk of hypertension. In eastern India including central Bihar, southern West Bengal, and some parts of Uttar Pradesh, there is a moderate to high risk of hypertension. The remaining regions show low to medium hypertension risk.

DISCUSSION

Our findings indicate significant correlations between physico-chemical groundwater parameters and hypertension, a concern given that approximately 85% of Indians rely on groundwater as their primary source of drinking water [10]. To our knowledge,

Table 6. Ordinary Least Squares (OLS) spatial lag and spatial error model to assess the association between the prevalence of hypertension and some selected background characteristics in India.

Indicators	OLS model		Spatial lag model		Spatial Error model	
	Coefficient	***p*-value**	**Coefficient**	***p*-value**	**Coefficient**	***p*-value**
Arsenic	2.212	0.001	2.143	0.001	2.296	0.001
Nitrate	−0.016	0.213	−0.001	0.182	0.001	0.117
Magnesium	0.022	0.024	0.009	0.036	0.007	0.019
Sulphate	0.105	0.001	0.102	0.003	0.107	0.001
Chloride	−0.002	0.129	−0.001	0.107	−0.001	0.090
Total Hardness	−0.006	0.223	−0.003	0.207	−0.003	0.203
Water pH	1.605	0.003	0.581	0.005	0.927	0.001
Non-vegetarian diet	0.012	0.136	0.013	0.102	0.070	0.006
AIC value	3171.8		2975.9		2960.2	
LAMBDA					**0.698**	0.001
RHO			**0.605**			
R-Square	**0.31**		**0.52**		**0.56**	
No of districts	593		593		593	

AIC Value (Akaike Information Criterion): A measure for model fit, with lower values indicating a better-fitting model.
Lambda: Represents spatial autocorrelation in the Spatial Error Model, indicating the degree of spatial dependence.
RHO (Spatial Lag Coefficient): The coefficient in the Spatial Lag Model, indicating the degree of spatial autocorrelation between neighbouring regions.
R-Square (Coefficient of Determination): The proportion of variance in the dependent variable explained by the model, with higher values indicating a better fit.

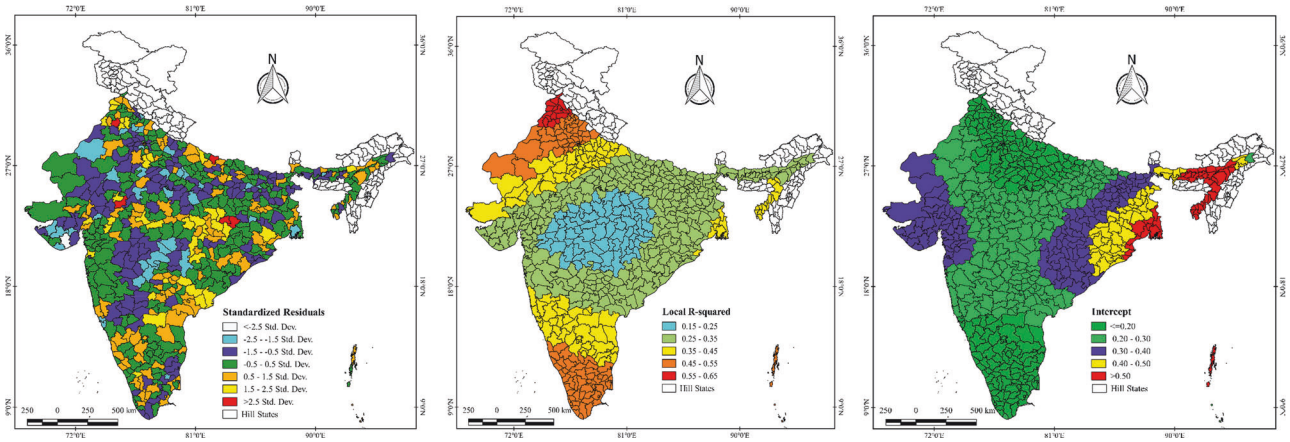

Fig. 7 Spatial pattern of standardized residuals, local R^2, and intercepts from the geographically weighted regression model.

this is the first study to examine the association of groundwater quality parameters and hypertension risk in India. Our study highlights important findings with highly relevant policy implications. Analysis of groundwater quality parameters, using data from 29,065 monitoring groundwater wells, revealed distinct spatial distribution patterns. Higher arsenic groundwater levels are primarily found in the northwestern, eastern, and northeastern regions of India, especially within the Ganga, Indus, and Brahmaputra river basins, while nitrate and fluoride are predominantly present in the northwestern and south-central regions. Individuals in areas with arsenic levels above 0.01 mg/L have 1.09 times higher odds of hypertension compared to those in areas with arsenic levels less than 0.01 mg/L. Similarly, nitrate levels above 45 mg/L are associated with 1.07 times higher odds of hypertension. Higher magnesium levels (>30 mg/L) increase the odds of hypertension by 1.03 times. Sulphate levels above 200 mg/L are associated with 1.16 times higher odds of hypertension. Elevated electrical conductivity (>300 µS/cm) increases the odds of hypertension by 1.06 times. For water pH, individuals in areas with pH above 8.5 have 2.12 times higher odds of hypertension compared to areas with pH ≤ 6.5.

High groundwater arsenic, nitrate and fluoride levels primarily originate from the geological formations and intensive agriculture. The alluvial deposits in the Indo-Gangetic Plain contain arsenic-rich pyrite that, when oxidized under certain hydrological conditions, releases arsenic into groundwater [41, 42]. Nitrate contamination is intensified by hydrological dynamics, which facilitate nitrate leaching into groundwater, particularly from agricultural regions with high fertilizer use [43]. The Green Revolution of late 1960s significantly transformed agriculture in northwestern India [44–47]. High-yield crop varieties, along with extensive use of chemical fertilizers and pesticides, has significantly increased agricultural productivity in India. This intensive agriculture and the use nitrogen-rich and sulphate-based fertilizers have contributed to high levels of nitrate and sulphate in groundwater, particularly in Punjab, Haryana, and Rajasthan. Sharma et al. [44] found that groundwater in southwestern Punjab has high electrical conductivity, total dissolved solids as well as fluoride, and nitrate levels, making it unsafe for drinking [44]. Studies in southern Punjab and Rajasthan also reported elevated salinity, fluoride, nitrate levels, and other contaminants [46, 47] Ahada and Suthar [46] found that intensive use of nitrogen-based fertilizer use increased nitrate levels in groundwater. Geological characteristics in northwestern India, such as alluvial plains and sandy soils, combined with erratic rainfall patterns, have further aggravated groundwater contamination [46–48]. Rapid urbanization and industrialization have further exacerbated groundwater contamination in northwestern India [49–51]. Industrial activities, including mining, manufacturing, and processing, have also contributed to groundwater contamination such as arsenic [48].

Our findings confirm t that elevated levels of arsenic, nitrate, magnesium and sulphate are associated with higher likelihood of hypertension. However, previous studies offer mixed conclusions on the association between arsenic exposure and hypertension risk [52–54]. Some studies found a strong correlation; others reported none or an inverse relationship [55]. Several studies showed that long-term exposure to arsenic [56–58] and nitrate [59] in drinking water can increase hypertension risk. Chronic exposure to arsenic in drinking water has been well studied as a risk factor for hypertension [53, 60, 61]. Arsenic disrupts vascular function by triggering oxidative stress, which alters gene expression, promotes inflammation, and reduces nitric oxide production [57, 62]. Studies in Bangladesh and parts of the U.S. show that populations exposed to arsenic-contaminated water have higher rates of hypertension [62–64]. A meta-analysis found that arsenic exposure showed a positive and approximately non-linear association with the risk of hypertension [60]. However, Chen et al. [52] in Bangladesh conducted a large population study ($n = 11{,}746$) and adjusted for confounders (e.g., age, sex, body mass index, tobacco smoking status), but they observed no association between arsenic exposure and hypertension risk. Guo et al. [52] reported a very high odds ratio (OR = 16.54) for the association between exposure from groundwater arsenic and hypertension [52] but the study had a small number of cases and did not adjust for confounders. Similarly, nitrate exposure can disrupt blood vessel function, raising hypertension risk [65]. Nitrate converts to nitric oxide, a vasodilator, but excessive exposure can produce harmful nitrogen compounds that damage blood cells [66]. While high nitrate intake has been associated with hypertension, its effects from drinking water are less studied [67, 68]. Ayub et al. [69] found significantly lower plasma nitrate levels in hypertensive patients compared to normotensive individuals [69]. However, other studies report the opposite trend, with lower hypertension risk observed in communities where drinking water contains higher nitrate levels [70]. Our study also revealed that higher calcium levels in the water (>75 mg/L) were associated with an 8% lower risk of hypertension. Calcium, in combination with sodium, potassium, and magnesium, helps maintain ionic balance in vascular membranes, promotes vasodilation and potentially reduces blood pressure [71–73]. These findings contradict a previous study, which found that calcium supplementation is a risk factor of increased blood pressure or hypertension [74]. However, optimal calcium levels play a crucial role in stabilizing vascular cell membranes, reducing calcium influx into cells, and consequently lessening vasoconstriction, which can help lower hypertension risk [75]. Daily intake recommendations generally range from 1000 to 1200 mg for adults, supporting

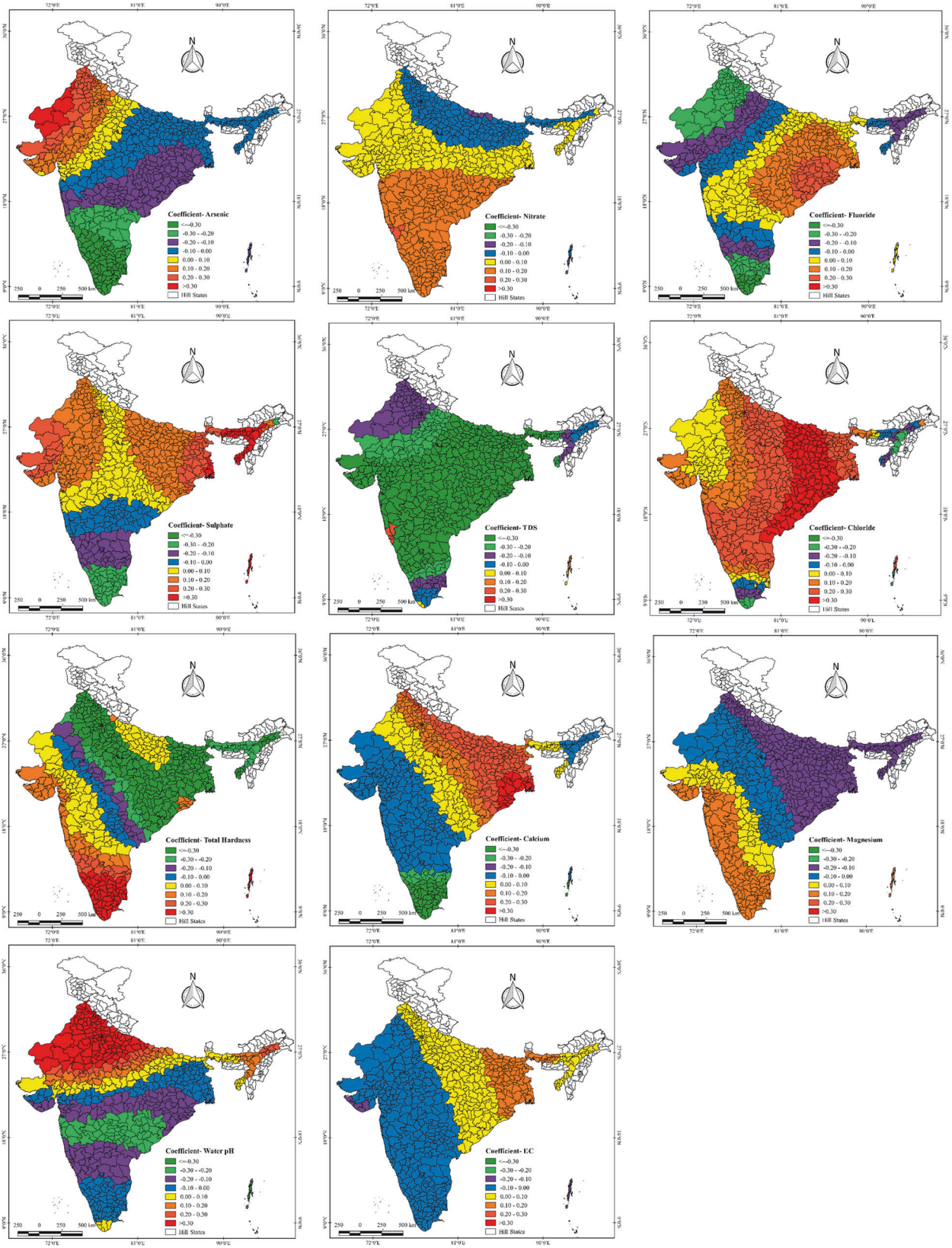

Fig. 8 Spatial pattern of coefficients of groundwater physicochemical properties and contaminants in association with hypertension in India – results from geographically weighted regression.

Table 7. Performance evaluation of machine learning models for hypertension risk prediction.

Model	Performance score		
	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	R^2 score
Artificial Neural Networks (ANN)	0.0236	0.0437	0.9038
Random Forest (RF)	0.0012	0.0077	0.9970
Extreme Gradient Boosting (XGBoost)	0.0977	0.1342	0.0930

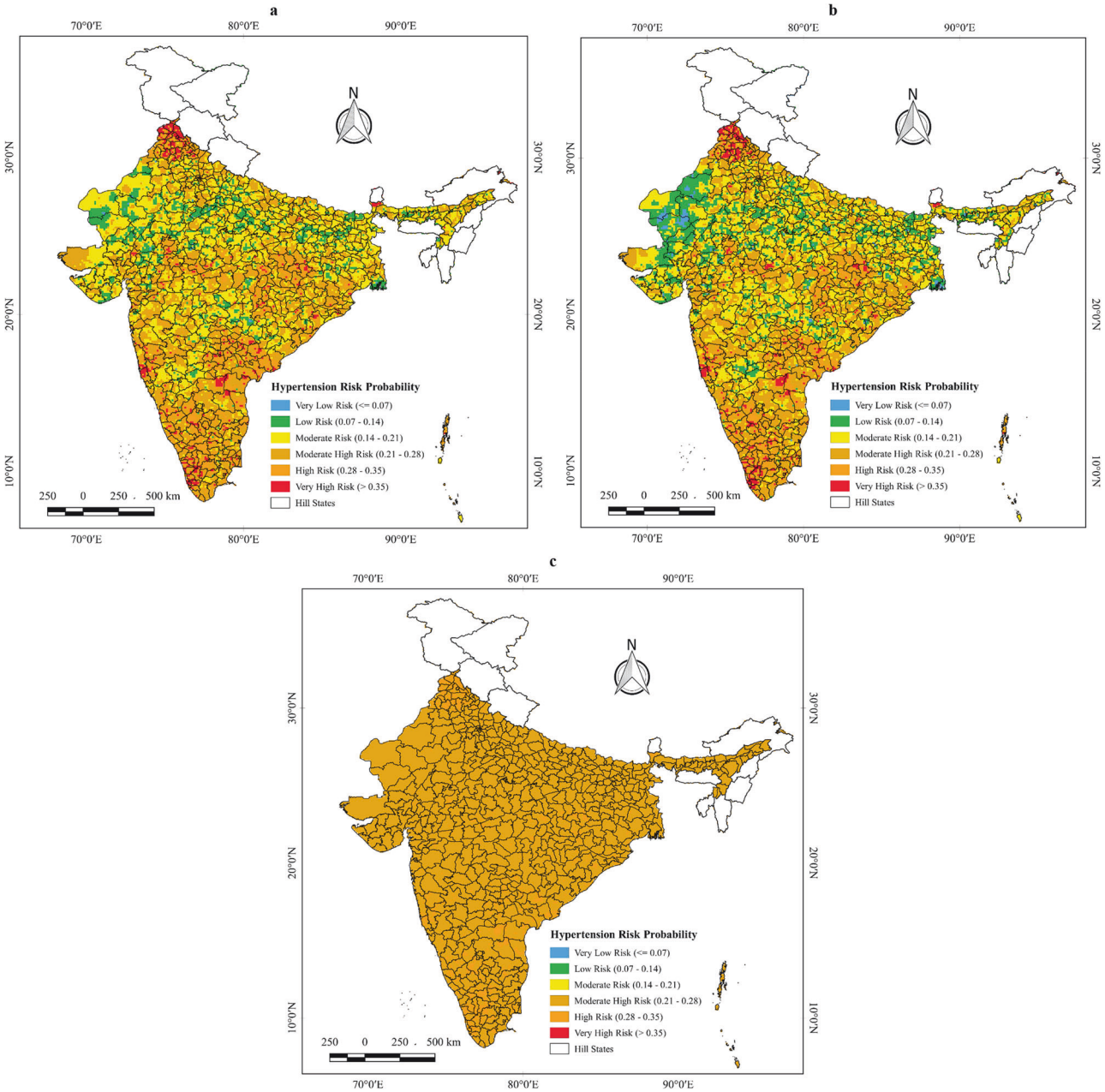

Fig. 9 Machine learning-based hypertension risk zonation in India using different predictive models. Machine learning-based hypertension risk zonation map: **a** artificial neural network, **b** random forest, and **c** extreme gradient boosting.

cardiovascular health and helping regulate blood pressure. This balance aids in maintaining vascular stability without the risks associated with excessive calcium levels [76]. Although calcium supplementation may not significantly affect blood pressure, the natural presence of calcium in drinking water could play a role in lowering the risk of hypertension.

Moreover, our study examines how demographic and socio-economic factors (e.g., age, rural/urban residence), as well as

dietary patterns affect hypertension risk. Our findings are consistent with previous research, highlighting that age, economic status, dietary habits, and geographical location contribute to hypertension risk [9, 77–82]. However, when comparing our findings to other literature, there are some discrepancies. Our study supports those of Ghosh and Kumar (Table 4) who also reported a significantly higher hypertension risk in older age groups [9]. Our study corroborates this finding, demonstrating a strikingly higher likelihood of hypertension in older age group compared to younger age groups. Our study also found that individuals of the higher and highest wealth groups face greater hypertension risk. This finding is consistent with other results demonstrating that the socioeconomic status is a major determinant of hypertension risk [78] Specifically, daily consumption of pulses, beans, and a vegetarian diet was linked to a lower risk of hypertension, while daily consumption of non-vegetarian foods showed an association with increased risk. Studies suggest that higher socioeconomic groups may have more access to unhealthy "fast food" options high in fats and meats, which could contribute to higher hypertension risks among wealthier individuals. This aligns with findings that greater fruit and vegetable intake is protective against hypertension [83]. Additionally, employees with higher nutrition knowledge were less likely to be hypertensive, highlighting the importance of dietary factors in hypertension prevention [81]. Moreover, our study found that rural residence had a lower likelihood of hypertension than urban residence. This finding aligns with the results of Dai et al. [80] who also found that rural residents were at a lower risk for hypertension [80]. However, Chaix et al. [79] found that the neighbourhood environment played a significant role in hypertension, with differences in prevalence observed across different geographical locations [79]. These studies did not consider the influence of environmental factors such as groundwater contaminants. By incorporating both environmental and sociodemographic factors, our study emphasizes the complex nature of hypertension etiology and thus the importance of interdisciplinary approaches in public health research.

Our study contributes to the broader discussion on the potential health risks associated with groundwater contamination and its link to hypertension in India. While our findings highlight associations between groundwater contaminants and hypertension risk, further validation through prospective studies or intervention-based research is necessary before making direct policy recommendations. Targeted interventions could help mitigate contamination levels, particularly in regions where arsenic, sulphate, pH, and nitrate exceed the BIS standard cut-off levels, such as within the Ganga, Indus, and Brahmaputra river basins. Future research should explore the effectiveness of water purification systems and alternative clean water sources to reduce contamination exposure and mitigate hypertension risk. The Random Forest model-generated hypertension risk map provides valuable insights that could aid in delineating high-risk zones across India, informing resource allocation for hypertension prevention and management strategies. Environmental monitoring, socioeconomic factors, and public health research should be combined to reduce hypertension risk in India. This integrated approach builds stronger evidence base and helps create effective strategies that address the multiple factors influencing hypertension in Indian communities.

This study has certain limitations. One important limitation is the potential for misclassification bias in defining hypertension. The definition included measured hypertension (systolic ≥140 mmHg or diastolic ≥90 mmHg) and self-reported use of antihypertensive medication. However, several antihypertensive drugs are also prescribed for conditions unrelated to hypertension, such as renal protection in diabetes, management of heart failure, and treatment of arrhythmias. This could have led to an overestimation of hypertension prevalence, as some individuals classified as hypertensive may have been taking these medications for other indications. Additionally, the accuracy of exposure classification is a key concern. Groundwater physicochemical and contaminant data were merged with NFHS-5 data at the cluster level. Individual water consumption patterns within a cluster may vary, making it challenging to precisely assess exposure levels. Another limitation is the exclusion of certain States and Union Territories due to missing groundwater data, which affects the generalizability of the findings. The absence of data from these regions may lead to an incomplete representation of groundwater quality and its potential health impacts across India. Future studies should consider primary data collection and longitudinal approaches to enhance exposure assessment and improve the robustness of the findings.

CONCLUSION

Our research sheds light on the complex relationship between groundwater quality and hypertension risk in India. Analysing extensive data from 29,065 monitoring wells in India, we identified distinct spatial patterns of contaminants, notably, arsenic, high pH, sulphate, and nitrate are concentrated in northern and northwestern regions of India, and this region was also found to have a high hypertension risk. We emphasize the importance of targeted interventions to mitigate contamination and reduce hypertension prevalence, alongside addressing socio-demographic factors like age and dietary patterns. Our findings underscore the need for interdisciplinary approaches, integrating environmental management, healthcare access, and lifestyle modifications to combat hypertension effectively. Utilizing advanced spatial analysis techniques like the Random Forest model, our study provides valuable insights for policymakers to delineate hypertension risk zones and allocate resources for prevention and management interventions across India.

DATA AVAILABILITY

The dataset analysed during the current study are available in the Demographic and Health Surveys (DHS) repository, https://www.dhsprogram.com/data/available-datasets.cfm.

REFERENCES

1. WHO. World Health Statistics. Global Health Estimates: Life expectancy and leading causes of death and disability. 2020. Available at: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates.
2. Mills KT, Stefanescu A, He J. The global epidemiology of hypertension. Nat Rev Nephrol. 2020;16:223–37.
3. Tomiyama H. Vascular function: A key player in hypertension. Hypertens Res. 2023;46:2145–58.
4. WHO. Global action plan for the prevention and control of noncommunicable diseases 2013-2020. World Health Organization, 2013.
5. WHO. Noncommunicable diseases progress monitor 2020. World Health Organization, 2020.
6. Laxmaiah A, Meshram II, Arlappa N, Balakrishna N, Rao KM, Reddy CG, et al. Socio-economic & demographic determinants of hypertension & knowledge, practices & risk behaviour of tribals in India. Indian J Med Res. 2015;141:697–708.
7. Ragavan RS, Ismail J, Evans RG, Srikanth VK, Kaye M, Joshi R, et al. Combining general and central measures of adiposity to identify risk of hypertension: a cross-sectional survey in rural India. Obes Res Clin Pr. 2023;17:249–56.
8. Indrapal M, Nagalla B, Varanasi B, Rachakulla H, Avula L. Socio-demographic factors, overweight/obesity and nutrients associated with hypertension among rural adults (≥18 years): Findings from National Nutrition Monitoring Bureau survey. Indian Heart J. 2022;74:382–90.
9. Ghosh S, Kumar M. Prevalence and associated risk factors of hypertension among persons aged 15–49 in India: a cross-sectional study. BMJ Open. 2019;9:e029714.
10. World Bank. India Groundwater: a Valuable but Diminishing Resource. 2012. https://www.worldbank.org/en/news/feature/2012/03/06/india-groundwater-critical-diminishing.
11. Adelodun B, Ajibade FO, Ighalo JO, Odey G, Ibrahim RG, Kareem KY, et al. Assessment of socioeconomic inequality based on virus-contaminated water usage in developing countries: a review. Environ Res. 2021;192:110309.

12. Asif M. The prevention and control the type-2 diabetes by changing lifestyle and dietary pattern. J Educ Health Promot. 2014;3:1.
13. Sarwar N, Ahmed T, Hossain A, Haque MM, Saha I, Sharmin KN. Association of dietary patterns with type 2 diabetes mellitus among Bangladeshi adults. Int J Nutr Sci. 2020;5:174–83.
14. Satija A, Hu FB, Bowen L, Bharathi AV, Vaz M, Prabhakaran D, et al. Dietary patterns in India and their association with obesity and central obesity. Public Health Nutr. 2015;18:3031–41.
15. Villegas R, Yang G, Gao Y-T, Cai H, Li H, Zheng W, et al. Dietary patterns are associated with lower incidence of type 2 diabetes in middle-aged women: the Shanghai Women's Health Study. Int J Epidemiol. 2010;39:889–99.
16. Bhatnagar RS, Padilla-Zakour OI. Plant-based dietary practices and socioeconomic factors that influence anemia in India. Nutrients. 2021;13:3538.
17. Agustina R, Nadiya K, Andini EA, Setianingsih AA, Sadariskar AA, Prafiantini E, et al. Associations of meal patterning, dietary quality and diversity with anemia and overweight-obesity among Indonesian school-going adolescent girls in West Java. PLoS One. 2020;15:e0231519.
18. Rammohan A, Awofeso N, Robitaille M-C. Addressing female iron-deficiency anaemia in India: is vegetarianism the major obstacle? Int Sch Res Not. 2012;2012:765476.
19. Ma J, Huang J, Zeng C, Zhong X, Zhang W, Zhang B, et al. Dietary Patterns and Association with Anemia in Children Aged 9–16 Years in Guangzhou, China: A Cross-Sectional Study. Nutrients. 2023;15:4133.
20. Baglietto L, Krishnan K, Severi G, Hodge A, Brinkman M, English DR, et al. Dietary patterns and risk of breast cancer. Br J Cancer. 2011;104:524–31.
21. Balasubramaniam SM, Rotti SB, Vivekanandam S. Risk factors of female breast carcinoma: a case control study at Puducherry. Indian J Cancer. 2013;50:65–70.
22. Kamath R, Mahajan KS, Ashok L, Sanal TS. A study on risk factors of breast cancer among patients attending the tertiary care hospital, in udupi district. Indian J Community Med. 2013;38:95–99.
23. Guha Mazumder D, Purkayastha I, Ghose A, Mistry G, Saha C, Nandy AK, et al. Hypertension in chronic arsenic exposure: A case control study in West Bengal. J Environ Sci Heal Part A. 2012;47:1514–20.
24. Khan JR, Awan N, Archie RJ, Sultana N, Muurlink O. The association between drinking water salinity and hypertension in coastal Bangladesh. Glob Heal J. 2020;4:153–8.
25. Manapurath RM, Anto RM, Pathak B, Malhotra S, Khanna P, Goel S. Diet and lifestyle risk factors associated with young adult hypertensives in India–Analysis of National Family Health Survey IV. J Fam Med Prim Care. 2022;11:5815–25.
26. EPA. National Primary Drinking Water Regulations. 2024. https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations.
27. WHO. Guidelines for drinking-water quality. 2024. https://iris.who.int/bitstream/handle/10665/375822/9789240088740-eng.pdf?sequence=1.
28. EU. European Drinking Water Directive. 2020. https://www.rehva.eu/fileadmin/user_upload/CELEX_32020L2184_EN_TXT.pdf.
29. BIS. BIS-Drinking Water Specifications (IS:10500-2012). 2012. https://cpcb.nic.in/wqm/BIS_Drinking_Water_Specification.pdf.
30. CCME. Canadian Water Quality Guidelines for the Protection of Aquatic Life. 2017. https://ccme.ca/en/res/wqimanualen.pdf.
31. Han J, Zhang L, Ye B, Gao S, Yao X, Shi X. The Standards for Drinking Water Quality of China (2022 Edition) Will Take Effect. China CDC Wkly. 2023;5:297.
32. NFHS-5. The DHS Program - India. 2022. https://dhsprogram.com/data/dataset_admin/index.cfm.
33. Garnero G, Godone D. Comparisons between different interpolation techniques. Int Arch Photogramm Remote Sens Spat Inf Sci. 2014;40:139–44.
34. Khan J, Mohanty SK. Spatial heterogeneity and correlates of child malnutrition in districts of India. BMC Public Health. 2018;18:1–13.
35. Khan J, Shil A, Prakash R. Exploring the spatial heterogeneity in different doses of vaccination coverage in India. PLoS One. 2018;13:e0207209.
36. Lu B, Charlton M, Harris P, Fotheringham AS. Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data. Int J Geogr Inf Sci. 2014;28:660–81.
37. Brunsdon C, Fotheringham S, Charlton M. Geographically weighted regression. J R Stat Soc Ser D. 1998;47:431–43.
38. Grossi E, Buscema M. Introduction to artificial neural networks. Eur J Gastroenterol Hepatol. 2007;19:1046–54.
39. Breiman L. Random forests. Mach Learn. 2001;45:5–32.
40. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: Association for Computing Machinery; 2016. pp. 785–94.
41. Saha D, Sahu S. A decade of investigations on groundwater arsenic contamination in Middle Ganga Plain, India. Environ Geochem Health. 2016;38:315–37.
42. Singh AK. Chemistry of arsenic in groundwater of Ganges–Brahmaputra river basin. Curr Sci. 2006;91:599–606.
43. Craswell E. Fertilizers and nitrate pollution of surface and ground water: an increasingly pervasive global problem. SN Appl Sci. 2021;3:518.
44. Sharma C, Mahajan A, Kumar Garg U. Fluoride and nitrate in groundwater of south-western Punjab, India—occurrence, distribution and statistical analysis. Desalin Water Treat. 2016;57:3928–39.
45. Tiwari KK, Krishan G, Prasad G, Mondal NC, Bhardwaj V. Evaluation of fluoride contamination in groundwater in a semi-arid region, Dausa District, Rajasthan, India. Groundw Sustain Dev. 2020;11:100465.
46. Ahada CPS, Suthar S. Groundwater nitrate contamination and associated human health risk assessment in southern districts of Punjab, India. Environ Sci Pollut Res. 2018;25:25336–47.
47. Coyte RM, Singh A, Furst KE, Mitch WA, Vengosh A. Co-occurrence of geogenic and anthropogenic contaminants in groundwater from Rajasthan, India. Sci Total Environ. 2019;688:1216–27.
48. Chetia M, Chatterjee S, Banerjee S, Nath MJ, Singh L, Srivastava RB, et al. Groundwater arsenic contamination in Brahmaputra river basin: a water quality assessment in Golaghat (Assam), India. Environ Monit Assess. 2011;173:371–85.
49. Ravish S, Setia B, Deswal S. Groundwater quality in urban and rural areas of north-eastern Haryana (India): a review. ISH J Hydraul Eng. 2021;27:224–34.
50. Wakode HB. Analysis of urban growth and assessment of impact of urbanization on water resources: a case study of Hyderabad, India [dissertation]. Aachen, Germany: Faculty of Georesources and Materials Engineering, RWTH Aachen University; 2016.
51. Bhardwaj SK, Sharma R, Aggarwal RK. Impact appraisal of industrialization on heavy metal contamination of Sirsa River located in the Shivalik Foothills of North Western Himalayas. Curr World Environ. 2019;14:245.
52. Guo JX, Hu L, Yand PZ, Tanabe K, Miyatalre M, Chen Y. Chronic arsenic poisoning in drinking water in Inner Mongolia and its associated health effects. J Environ Sci Heal Part A. 2007;42:1853–8.
53. Chen Y, Factor-Litvak P, Howe GR, Graziano JH, Brandt-Rauf P, Parvez F, et al. Arsenic exposure from drinking water, dietary intakes of B vitamins and folate, and risk of high blood pressure in Bangladesh: a population-based, cross-sectional study. Am J Epidemiol. 2007;165:541–52.
54. Dastgiri S, Mosaferi M, Fizi MAH, Olfati N, Zolali S, Pouladi N, et al. Arsenic exposure, dermatological lesions, hypertension, and chromosomal abnormalities among people in a rural community of northwest Iran. J Health Popul Nutr. 2010;28:14.
55. Guo J, Cao W, Lang G, Sun Q, Nan T, Li X, et al. Worldwide distribution, health risk, treatment technology, and development tendency of geogenic high-arsenic groundwater. Water. 2024;16:478.
56. Abhyankar LN, Jones MR, Guallar E, Navas-Acien A. Arsenic exposure and hypertension: a systematic review. Environ Health Perspect. 2012;120:494–500.
57. Balarastaghi S, Rezaee R, Hayes AW, Yarmohammadi F, Karimi G. Mechanisms of arsenic exposure-induced hypertension and atherosclerosis: an updated overview. Biol Trace Elem Res. 2023;201:98–113.
58. Xu J, Engel LS, Rhoden J, Jackson IIWB, Kwok RK, Sandler DP. The association between blood metals and hypertension in the GuLF study. Environ Res. 2021;202:111734.
59. Xeni C, Oliva R, Jahan F, Romaina I, Naser AM, Rahman M, et al. Epidemiological evidence on drinking water salinity and blood pressure: a scoping review. Environ Res Heal. 2023;1:35006.
60. Zhao J, Li A, Mei Y, Zhou Q, Li Y, Li K, et al. The association of arsenic exposure with hypertension and blood pressure: a systematic review and dose–response meta-analysis. Environ Pollut. 2021;289:117914.
61. Jensen GE, Hansen ML. Occupational arsenic exposure and glycosylated haemoglobin. Analyst. 1998;123:77–80.
62. Khatun M, Haque N, Siddique AE, Wahed AS, Islam MS, Khan S, et al. Arsenic exposure-related hypertension in bangladesh and reduced circulating nitric oxide bioavailability. Environ Health Perspect. 2024;132:47003.
63. Islam MR, Khan I, Attia J, Hassan SMN, McEvoy M, D'Este C, et al. Association between hypertension and chronic arsenic exposure in drinking water: a cross-sectional study in Bangladesh. Int J Environ Res Public Health. 2012;9:4522–36.
64. Kaufman JA, Mattison C, Fretts AM, Umans JG, Cole SA, Voruganti VS, et al. Arsenic, blood pressure, and hypertension in the Strong Heart Family Study. Environ Res. 2021;195:110864.
65. Ward MH, Jones RR, Brender JD, De Kok TM, Weyer PJ, Nolan BT, et al. Drinking water nitrate and human health: an updated review. Int J Environ Res Public Health. 2018;15:1557.
66. Tejero J, Shiva S, Gladwin MT. Sources of vascular nitric oxide and reactive oxygen species and their regulation. Physiol Rev. 2019;99:311–79.
67. Levin R, Villanueva CM, Beene D, Cradock AL, Donat-Vargas C, Lewis J, et al. US drinking water quality: exposure risk profiles for seven legacy and emerging contaminants. J Expo Sci Environ Epidemiol. 2024;34:3–22.
68. Rahman A, Mondal NC, Tiwari KK. Anthropogenic nitrate in groundwater and its health risks in the view of background concentration in a semi arid area of Rajasthan, India. Sci Rep. 2021;11:1–13.

69. Ayub T, Khan SN, Ayub SG, Dar R, Andrabi KI. Reduced nitrate level in individuals with hypertension and diabetes. J Cardiovasc Dis Res. 2011;2:172–6.
70. Malberg JW, Savage EP, Osteryoung J. Nitrates in drinking water and the early onset of hypertension. Environ Pollut. 1978;15:155–60.
71. Wabo TMC, Wu X, Sun C, Boah M, Nkondjock VRN, Cheruiyot JK, et al. Association of dietary calcium, magnesium, sodium, and potassium intake and hypertension: a study on an 8-year dietary intake data from the National Health and Nutrition Examination Survey. Nutr Res Pr. 2022;16:74–93.
72. Melaku L, Elias B. The Physiological Mechanism of Extracellular CalciumSensing Receptor Action in the Regulation of Vascular Tone and Blood Pressure. Biomed J Sci Tech Res. 2023;51:43156–65.
73. Ma J, Li Y, Yang X, Liu K, Zhang X, Zuo X, et al. Signaling pathways in vascular function and hypertension: molecular mechanisms and therapeutic interventions. Sig Transduct Target Ther. 2023;8:168.
74. Whelton PK, Appel L, Charleston J, Dalcin AT, Ewart C, Fried L, et al. The effects of nonpharmacologic interventions on blood pressure of persons with high normal levels: results of the Trials of Hypertension Prevention, phase I. JAMA. 1992;267:1213–20.
75. Das UN. Nutritional factors in the pathobiology of human essential hypertension. Nutrition. 2001;17:337–46.
76. Harvard Medical School. How much calcium do you really need? 2022. https://www.health.harvard.edu/staying-healthy/how-much-calcium-do-you-really-need.
77. Abba MS, Nduka CU, Anjorin S, Mohamed SF, Agogo E, Uthman OA. Influence of contextual socioeconomic position on hypertension risk in low-and middle-income countries: disentangling context from composition. BMC Public Health. 2021;21:1–13.
78. Ghosh PK, Harun MGD, Shanta IS, Islam A, Jannat KKE, Mannan H. Prevalence and determinants of hypertension among older adults: A comparative analysis of the 6th and 8th national health surveys of Bangladesh. PLoS One. 2023;18:e0292989.
79. Chaix B, Bean K, Leal C, Thomas F, Havard S, Evans D, et al. Individual/neighborhood social factors and blood pressure in the RECORD Cohort Study: which risk factors explain the associations?. Hypertension. 2010;55:769–75.
80. Dai B, Addai-Dansoh S, Nutakor JA, Osei-Kwakye J, Larnyo E, Oppong S, et al. The prevalence of hypertension and its associated risk factors among older adults in Ghana. Front Cardiovasc Med. 2022;9:990616.
81. Geaney F, Fitzgerald S, Harrington JM, Kelly C, Greiner BA, Perry IJ. Nutrition knowledge, diet quality and hypertension in a working population. Prev Med Rep. 2015;2:105–13.
82. Rajkumar E, Romate J. Behavioural risk factors, hypertension knowledge, and hypertension in rural India. Int J Hypertens. 2020;2020:8108202.
83. Rao ND, Min J, DeFries R, Ghosh-Jerath S, Valin H, Fanzo J. Healthy, affordable and climate-friendly diets in India. Glob Environ Chang. 2018;49:154–65.

ACKNOWLEDGEMENTS

This study is part of Sourav Biswas's Ph.D. research, and he gratefully acknowledges the support and academic environment provided by the International Institute for Population Sciences (IIPS), Mumbai, which made this work possible. This paper was presented at the 35th International Geographical Congress (IGC) 2024, held in Dublin, Ireland. The authors acknowledge the financial support provided by the Anusandhan National Research Foundation (ANRF), Government of India, which enabled attendance at the conference through travel funding. The authors are also thankful to the reviewers for their excellent comments and constructive suggestions that helped improve the quality of this manuscript. We extend our sincere thanks to the editor for the guidance throughout the review process.

AUTHOR CONTRIBUTIONS

Sourav Biswas is the guarantor of this work. Sourav Biswas: Conceptualisation, methodology, formal analysis, resource acquisition, preparation of maps and tables, writing—original draft, writing—review and editing. Aparajita Chattopadhyay: Conceptualisation, resource acquisition, writing—original draft, writing—review and editing, and supervision. Kathrin Schilling: Conceptualisation, writing—original draft, writing—review and editing, and supervision. Ayushi Das: Methodology, formal analysis, writing—review and editing.

FUNDING

This study did not receive any specific grant from any funding agencies.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study is based on secondary analysis of publicly available, anonymized data from the National Family Health Survey (NFHS-5), which is the Indian version of the Demographic and Health Survey (DHS). The survey protocol, including procedures for data collection and informed consent, received ethical approval from the Ministry of Health and Family Welfare (MoHFW), Government of India. All methods were performed in accordance with relevant institutional and national guidelines and regulations. The NFHS-5 data were collected following strict ethical standards, including obtaining verbal and written informed consent from all participants prior to their inclusion in the survey. In cases involving minors, informed consent was obtained from a parent or legal guardian. As the study involved secondary, de-identified data, no additional ethical approval or anonymization was required. The dataset is publicly available and can be accessed at: http://www.dhsprogram.com/data/available-datasets.cfm. No identifiable images or personal details of participants were used; hence, consent for publication is not applicable.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Aparajita Chattopadhyay.