Is Groundwater Contamination Contributing to Women's Anemia in India? Using Machine Learning Approach for Risk Zone Predictions

Sourav Biswas, Aparajita Chattopadhyay, Aditi B Prasad

Introduction:

Groundwater plays a vital role in India's water supply, with approximately 85% of the population relying on it for drinking purposes (WHO, 2012). However, rapid population growth, urbanization, and expanding agricultural activities have significantly increased groundwater extraction, exceeding sustainable levels. Currently, more than 60% of the country's groundwater is being extracted, leading to a critical decline in water quality (UNESCO, 2022). Contaminants such as fluoride, salinity, arsenic, and heavy metals are commonly found in groundwater, making it unsafe for consumption in around 27% of India's geographical areas. Despite this, many people continue to drink unfiltered groundwater, posing a serious public health risk (Biswas et al., 2024). Contaminated groundwater has been linked to numerous health problems, including cancer, skeletal and dental fluorosis, blue baby syndrome, cardiovascular disease, kidney damage, and gastrointestinal disorders (Bini & Wahsha, 2014). Previous research has primarily focused on socio-demographic factors, nutrition, and dietary habits as key contributors to anemia (Bentley & Griffiths, 2003; Let et al., 2024; Sharif et al., 2023), but the potential impact of drinking water contamination has received little attention. To date, no comprehensive investigation has explored the association between water contaminants and anemia in the Indian context. Addressing anemia in women of reproductive age is crucial for improving overall public health, enhancing maternal well-being, and contributing to economic growth. This study seeks to fill this critical research gap by examining the relationship between water contamination and anemia among reproductive-aged women. Furthermore, machine learning techniques were employed to predict high-risk zones for anemia based on identified environmental and health predictors across different regions of India.

Data:

We utilized data from two distinct sources for this study. Groundwater data, collected between 2019 and 2021, was obtained from the Central Groundwater Board (CGWB), which sampled 29,065 locations. The acceptable limits for drinking water quality were based on the Bureau of Indian Standards (BIS), and Table 1 compares these limits with those set by other international organizations. Anemia and socio-demographic data were sourced from India's National Family Health Survey (NFHS-5), conducted during the same period. The analysis includes 581,286 individuals, with Himalayan states (Jammu and Kashmir, Himachal Pradesh, Uttarakhand, Sikkim, Meghalaya, Arunachal Pradesh, Mizoram, and Nagaland) excluded due to the lack of CGWB water samples in these regions. To integrate the groundwater contamination data with NFHS-5 data, we employed the cluster shapefile from the NFHS-5 database. Using this shapefile, we extracted contaminant values from the groundwater raster maps and linked them with individual records by merging the cluster IDs with the water data. For geospatial analysis, Zonal Statistics was applied to calculate the mean values of water contaminants at the district level across India.

Table-1: Acceptable limits for drinking water quality parameters from BIS, WHO, EU, China, and U.S. EPA.

Parameters	BIS (India)	WHO	EU	China	US EPA
Arsenic (as As)	0.01 mg/L	0.01 mg/L	0.01 mg/L	0.01 mg/L	0.01 mg/L
Nitrate (as NO_3)	45 mg/L	50 mg/L	50 mg/L	20 mg/L	10 mg/L
Fluoride (as F^-)	1.0 mg/L	1.5 mg/L	1.5 mg/L	1.0 mg/L	4.0 mg/L
Sulphate (as $SO4^{2-}$)	200 mg/L	500 mg/L	250 mg/L	250 mg/L	250 mg/L
Total Dissolved Solids (TDS)	500 mg/L	600 mg/L	No limit specified	1000 mg/L	500 mg/L
Chloride (as Cl^-)	250 mg/L	No guideline	250 mg/L	250 mg/L	250 mg/L
Total Hardness (as $CaCO_3$)	300 mg/L	No guideline	No limit specified	450 mg/L	No limit specified
Calcium (as Ca^{2+})	75 mg/L	No guideline	No limit specified	No limit specified	No limit specified
Magnesium (as Mg^{2+})	30 mg/L	No guideline	No limit specified	No limit specified	No limit specified
Iron (as Fe)	0.3 mg/L	0.3 mg/L	0.2 mg/L	0.3 mg/L	0.3 mg/L
pH	6.5 - 8.5	6.5 - 8.5	6.5 - 9.5	6.5 - 8.5	6.5 - 8.5
Electrical Conductivity (EC)	300 µS/cm	1500 µS/cm	2500 µS/cm	Not specified	No limit specified

Note: For analysis purposes, we follow the Bureau of Indian Standard (e.g. arsenic levels are considered normal when they are less than or equal to 0.01 mg/L, while levels exceeding 0.01 mg/L are deemed unsafe).

Variable description:

Anemia, the outcome variable in this study, is derived from the individual women's data file (aged 15-49) in the NFHS-5 dataset. It is categorized into non-anemic (coded as 0) and anemic (coded as 1). The primary explanatory variables include arsenic, nitrate, fluoride, sulphate, total dissolved solids, chloride, total hardness, calcium, magnesium, iron, water pH, and electrical conductivity, all sourced from the CGWB. Each parameter was categorized as either within the acceptable limit (coded as 0) or above the acceptable limit (coded as 1), based on BIS standards. Additional predictor variables potentially influencing anemia were adjusted for in the analysis. These include body mass index (underweight, normal, overweight/obese), education (no education, primary, secondary, higher), religion (Hindu, Muslim, Christian, others), wealth (poorest, poorer, middle, richer, richest), food and dietary habits (pulses, green vegetables, fruits, chicken/meat: Never/Occasionally, Weekly, Daily), and region (north, central, east, northeast, west, south).

Methods:

A Chi-square-based bivariate analysis was performed to assess the relationship between women's anemia and various socio-environmental factors. To further examine the connection between water contaminants and anemia, we applied adjusted logistic regression to the dataset, accounting for potential confounding variables. Spatial associations between anemia and water contaminants at the district level were analyzed using Bivariate Local Moran's I statistics, along with LISA cluster maps. Once the key groundwater contaminants linked to anemia were identified, we predicted high-risk zones for anemia using three supervised machine learning models: Artificial Neural Networks (ANN), Random Forest (RF), and Extreme Gradient Boosting (XGBoost).

Result:

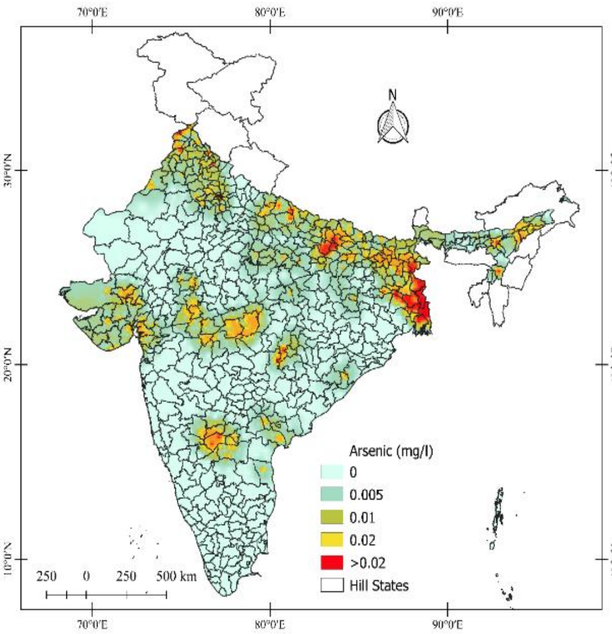
Figure 1 shows the spatial distribution of various groundwater contaminants across India, indicating that high arsenic levels are concentrated in the east, northeast, and northern regions, while other contaminants are more prevalent in the western and south-central areas (specific figures not shown). The bivariate analysis (not given here) of women aged 15-49 showed that 57% were anemic, with significant links between groundwater contaminants and anemia. For instance, 60% of women in areas with high arsenic levels (>0.01 mg/L) were anemic, compared to 56% in regions with arsenic within safe limits. Similarly, 59% of women in areas with higher fluoride levels (>1.0 mg/L) were anemic, compared to 57% where fluoride was below 1.0 mg/L. Nitrate (>45 mg/L) and sulfate (>200 mg/L) showed similar patterns to fluoride, while TDS levels mirrored those of arsenic.

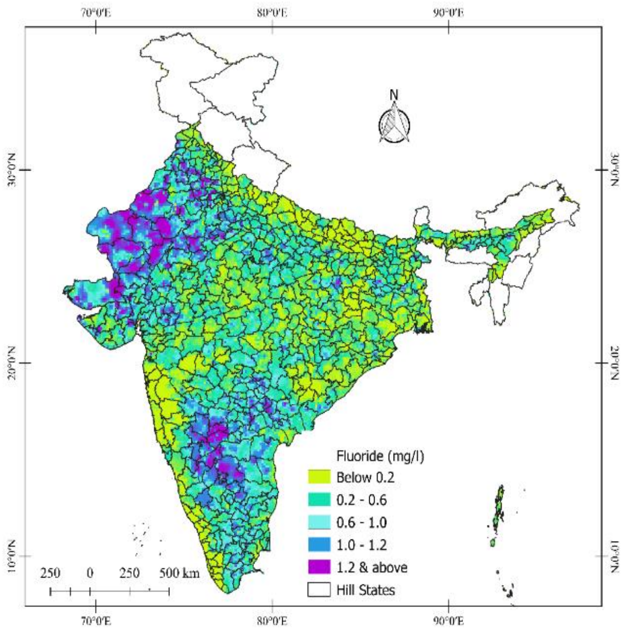
The adjusted logistic regression results (Table 2) showed that women exposed to high levels of groundwater contaminants had increased odds of anemia, with the likelihood rising by 32% for pH, 13% for TDS, 11% for total hardness, 10% for arsenic, 9% for iron, 7% for nitrate, 6% for sulphate, 5% for electrical conductivity, and 3% for fluoride compared to areas where these values were within acceptable BIS limits.

The Bivariate LISA cluster map and Moran's I statistics (figure-2) (0.19) indicated that high arsenic levels were a significant risk factor for anemia, particularly in eastern and northeastern India.

In predicting anemia risk zones using machine learning, the Random Forest (RF) model was the best performer, achieving an R^2 of 0.9912, MAE of 0.0006, and MSE of 0.0072. The Artificial Neural Network (ANN) model followed, with an R^2 of 0.9495, MAE of 0.0178, and MSE of 0.0342. XGBoost had the lowest accuracy, with an R^2 of 0.3693, MAE of 0.1009, and MSE of 0.121. The RF model identified higher anemia risks in northern, eastern, northeastern, and south-central regions of India.

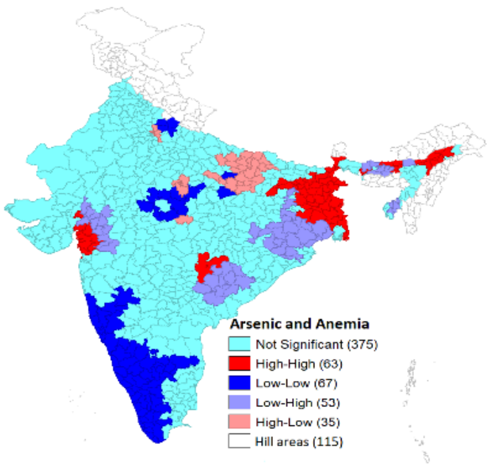
Figure 1: Spatial distribution of arsenic and fluoride across India, 2019-2021 (other contaminant figures not shown).



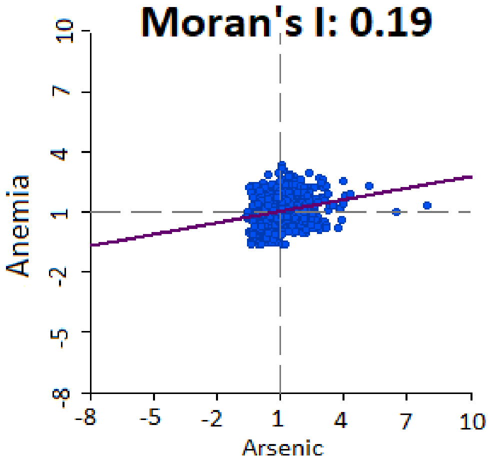


A: Arsenic B: Fluoride

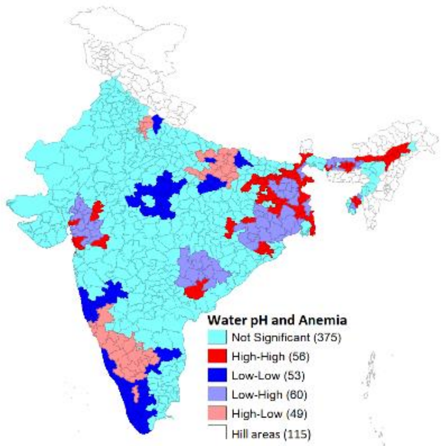
Figure 2 (a-d): Bivariate LISA cluster maps and scatter plots depicting geographic clusters of (a-b) arsenic and anemia, and (c-d) water pH and anemia in India.



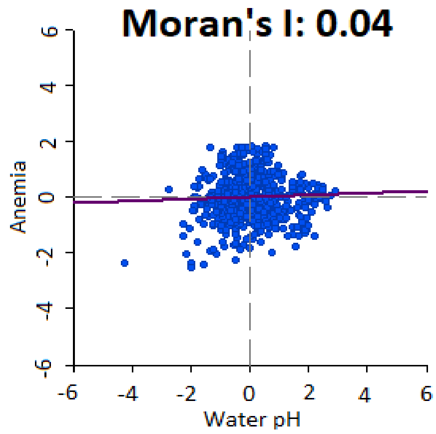
A: Arsenic and Anemia



B: Arsenic and Anemia



C: Water pH and Anemia



D: Water pH and Anemia

Discussion:

The primary hypothesis of this study was that carcinogenic groundwater contaminants such as Arsenic, Nitrate, Fluoride, and Sulphate, present above acceptable limits, are associated with anemia among reproductive-aged women in India. This hypothesis was supported by the findings. Logistic regression results revealed significant associations between groundwater contaminants and anemia. Women exposed to high Arsenic levels (>0.01 mg/L) had a 10% higher risk of anemia, while elevated Nitrate levels (>45 mg/L) were linked with a 7% increased risk. Similarly, high Fluoride and Sulphate levels were associated with 3% and 6% higher odds of anemia after adjusting for confounding factors such as socio-demographic characteristics and dietary habits. Machine learning models further confirmed these associations, with the Random Forest (RF) model showing the highest predictive accuracy (MAE: 0.0006, MSE: 0.0072, R^2: 0.9912). RF spatial analysis highlighted northern, eastern, and northeastern regions as high-risk zones for anemia. Chronic exposure to arsenic, prevalent in areas such as West Bengal, Bihar, Assam damages bone marrow and disrupts red blood cell production, leading to anemia. Arsenic poisoning can also cause hemolysis and interfere with iron metabolism, preventing the body from using iron efficiently (Shander et al., 2009). On the other hand, Nitrates are converted into nitrites in the body, which oxidize iron in haemoglobin, reducing its oxygen-carrying capacity and leading to anemia (Kim-Shapiro et al., 2006). Furthermore, Iron found in groundwater is inorganic and poorly absorbable, making it less bioavailable to the body and contributing to iron-deficiency anemia (Ghosh et al., 2020). These findings highlight groundwater contamination's significant role in anemia prevalence, underscoring the need for targeted public health interventions.

Table 2: Adjusted Odds Ratios (AOR) shows the relationship between Groundwater contaminants and Anemia in India.

Selected Characteristics	AOR [95 % CI]
n= 581286	**Women's Anemia**
Arsenic (as As)	
≤0.01 mg/L®	Ref
>0.01 mg/L	1.10***[1.08,1.11]
Nitrate (as NO3)	
≤45 mg/L®	Ref
>45 mg/L	1.07***[1.06,1.09]
Fluoride (as F−)	
≤1 mg/L®	Ref
>1 mg/L	1.03**[1.00,1.06]
Sulphate (as $SO4^{2-}$)	
≤200 mg/L®	Ref
>200 mg/L	1.06**[1.01,1.11]
Total Dissolved Solids (TDS)	
≤500 mg/L®	Ref
>500 mg/L	1.13***[1.11,1.14]
Chloride (as Cl−)	
≤250 mg/L®	Ref
>250 mg/L	0.89*[0.79,0.99]
Total Hardness (as CaCO3)	
≤300 mg/L®	Ref
>300 mg/L	1.11***[1.08,1.15]
Calcium (as Ca^{2+})	
≤75 mg/L®	Ref
>75 mg/L	0.96[0.88,1.04]
Magnesium (as Mg^{2+})	
≤30 mg/ L®	Ref
>30 mg/L	0.92*[0.86,0.98]
Iron (as Fe)	
≤0.3 mg/L®	Ref
>0.3 mg/L	1.09***[1.07,1.11]
pH	
<6.5®	Ref
6.5 - 8.5	1.03[0.94,1.12]
>8.5	1.32***[1.26,1.37]
Electrical Conductivity (EC)	
≤300 µS/cm®	Ref
>300 µS/cm	1.05**[1.01,1.09]

Note: *** significant at 1%; ** significant at 5% and * significant at 10%.

Conclusion

As India faces a growing water pollution crisis, addressing groundwater contamination with Arsenic, Fluoride, Nitrate, and other harmful substances is essential for improving both general health and reproductive outcomes. Ensuring access to safe drinking water through filtration systems and implementing stricter water quality regulations can significantly reduce anemia and other health risks. Policymakers should prioritize intervention in the high-risk zones identified by machine learning methods, ensuring targeted measures in these areas to combat anemia effectively.

References:

Bentley, M. E., & Griffiths, P. L. (2003). The burden of anemia among women in India. *European Journal of Clinical Nutrition*, *57*(1), 52–60.

Bini, C., & Wahsha, M. (2014). Potentially harmful elements and human health. *PHEs, Environment and Human Health: Potentially Harmful Elements in the Environment and the Impact on Human Health*, 401–463.

Biswas, S., Chattopadhyay, A., Shaw, S., & Hoffmann, R. (2024). Assessing groundwater quality and its association with child undernutrition in India. *Science of the Total Environment*, *943*, 173732.

Ghosh, G. C., Khan, M. J. H., Chakraborty, T. K., Zaman, S., Kabir, A. H. M. E., & Tanaka, H. (2020). Human health risk assessment of elevated and variable iron and manganese intake with arsenic-safe groundwater in Jashore, Bangladesh. *Scientific Reports*, *10*(1), 5206.

Kim-Shapiro, D. B., Schechter, A. N., & Gladwin, M. T. (2006). Unraveling the reactions of nitric oxide, nitrite, and hemoglobin in physiology and therapeutics. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *26*(4), 697–705.

Let, S., Tiwari, S., Singh, A., & Chakrabarty, M. (2024). Prevalence and determinants of anaemia among women of reproductive age in Aspirational Districts of India: an analysis of NFHS 4 and NFHS 5 data. *BMC Public Health*, *24*(1), 437.

NFHS-5. (2022). *National Family Health Survey (NFHS-5)- 2019-20*. International Institute for Population Sciences. https://www.nfhs.in/nfhsuser/nfhs5.php

Shander, A., Cappellini, M. D., & Goodnough, L. T. (2009). Iron overload and toxicity: the hidden risk of multiple blood transfusions. *Vox Sanguinis*, *97*(3), 185–197.

Sharif, N., Das, B., & Alam, A. (2023). Prevalence of anemia among reproductive women in different social group in India: cross-sectional study using nationally representative data. *Plos One*, *18*(2), e0281015.

UNESCO. (2022). *The United Nations World Water Development Report 2022*. https://unesdoc.unesco.org/ark:/48223/pf0000380721

WHO. (2012). *India Groundwater: a Valuable but Diminishing Resource*. World Bank Group. https://www.worldbank.org/en/news/feature/2012/03/06/india-groundwater-critical-diminishing#:~:text=India is the largest user of groundwater in,of drinking water supplies are dependent on groundwater.