JOWRM

Journal of Water Resources Management

Journal homepage: https://journal.water.gov.my

e-ISSN: 2811-3578

Health Risk Assessment of Physicochemical and Heavy Metal Exposures of Groundwater in Linggi River Basin, Malaysia with Machine Learning Approach

Nurul Amirah Najwa Azlan[1,a], Noorain Mohd Isa[1,b*], Looi Ley Juen[1,c], Ahmad Zaharin Aris[1,d]

[1]Universiti Putra Malaysia, Serdang, 43400, Malaysia

Email: [a]214638@student.upm.edu.my, [b,]*mi.noorain@upm.edu.my, [c]leyjuenlooi@upm.edu.my, [d]zaharin@upm.edu.my,

Received 02 Mac 2026;
Accepted 23 Jun 2026;
Available online 30 Jun 2026

Abstract: Groundwater is an increasingly important water source in Malaysia due to climate variability, population growth, and rising demand. However, contamination by heavy metals poses significant risks to public health and water sustainability. This study aimed to assess the physicochemical quality of groundwater in the Linggi River Basin and evaluate health risks associated with heavy metal exposure. A total of 11 sampling wells were selected between 2022 and 2024, representing diverse land-use activities. Major heavy metals (Cd, Cu, Fe, Pb, Mn, Ni, Zn) were analyzed using ICP-MS and compared against World Health Organization standards. Statistical analyses and machine learning models, specifically Random Forest (RF) and Multiple Linear Regression (MLR), were applied to predict metal concentrations, while Shapley Additive Explanations (SHAP) identified key geochemical drivers. Health risks were quantified using Hazard Quotient (HQ), Hazard Index (HI), and Excess Cancer Risk (ECR) for adults and children via oral and dermal pathways. Results revealed severe contamination, with Fe (mean 164.53 mg/L) and Mn (mean 6.93 mg/L) far exceeding WHO limits, while Pb and Cd also surpassed permissible levels at several sites. RF outperformed MLR in predictive accuracy, and SHAP confirmed the influence of oxidation-reduction potential (ORP), HCO_3, and temperature on metal mobility. Health risk assessment showed oral ingestion as the dominant exposure pathway, with children more vulnerable than adults. Cd emerged as the primary carcinogenic agent, contributing over 90% of cumulative cancer risk. Agricultural and mixed-use zones exhibited the highest risk indices, reflecting combined geogenic enrichment and anthropogenic inputs. In conclusion, this study highlights the urgent need for groundwater monitoring, pollution control, and child-focused health interventions in high-risk zones. By integrating machine learning, health risk assessment, and GIS mapping, the research provides a scalable framework for early detection, hotspot identification, and improved groundwater governance in Malaysia.

Keywords: Groundwater quality; Heavy metal contamination; Random Forest (RF); Multiple Linear Regression (MLR); GIS mapping; SHAP (Shapley Additive Explanations)

1. Introduction

Around 36% of the world's drinking water and 40% of its irrigation demands are met by groundwater, which is an essential source of freshwater for drinking, irrigation, and industrial usage [1]. While surface water is still the main source of water in Malaysia, groundwater is becoming more significant in rural and semi-urban areas with inadequate infrastructure [2]. However, contamination by heavy metals such as Fe, Mn, Pb, and Cd poses serious risks to public health. These metals originate from industrial discharge, agricultural runoff, and geogenic processes, and are associated with neurotoxicity, carcinogenicity, and organ damage [3].

The Linggi River Basin in Negeri Sembilan is a key aquifer system supporting domestic and agricultural needs. Limited monitoring has raised concerns about contamination, particularly in agro-industrial zones where heavy metal concentrations often exceed WHO guidelines [4]. To address these gaps, this study integrates machine learning, health risk assessment, and GIS mapping to evaluate groundwater quality and its associated health risks. The objectives are to assess physicochemical parameters, predict heavy metal concentrations, identify geochemical drivers, and quantify non-carcinogenic and carcinogenic risks for adults and children.

2. Materials and Methodology

2.1. Study Area

The Linggi River Basin is located in Negeri Sembilan, Malaysia, and represents one of the country's important shallow aquifer systems. Rainfall and riverbank filtration are the primary recharge sources, while evaporation and agricultural withdrawals contribute to losses[5]. The regional context of the study area is shown in Figure 1, which situates

**Corresponding author: mi.noorain@upm.edu.my*
2026 JPS Publishing. All right reserved.

the Linggi River Basin within Southeast Asia, Malaysia, and Negeri Sembilan.

Geologically, the basin is underlain by a mixture of alluvial deposits, sedimentary formations, and weathered bedrock. Aquifers consist of sand, gravel, and fractured rock units, which provide moderate to high groundwater potential depending on depth and permeability. Soil series distribution further influences groundwater quality, with clay-rich soils limiting infiltration and sandy soils enhancing recharge [5].

Land use within the Linggi River Basin is diverse, including agricultural zones, industrial estates, and mixed residential areas [6]. Fertilisers application contributes to Cd and NO_3 enrichment in groundwater [7]. Industrial operations such as electroplating, textile dyeing, and battery manufacturing are major sources of lead, nickel, and copper contamination [2]. Residential and mixed-use areas also exert pressure on groundwater through domestic wastewater discharge and landfill leachate [3]. The combination of geogenic enrichment and anthropogenic inputs makes the basin highly vulnerable to heavy metal contamination, necessitating integrated monitoring and risk assessment.

2.2. Sample Collection, Preparation and Analysis

Groundwater samples were collected from shallow aquifers at 11 designated sampling sites (SS1–SS11) across the Linggi River Basin. These sites were selected to represent diverse hydrogeological and land-use conditions, including agricultural, industrial, and mixed residential zones. Sampling locations were chosen based on accessibility, borehole depth, and proximity to potential contamination sources, ensuring comprehensive coverage of the basin.

Standard procedures recommended by the American Public Health Association were followed [8]. Each well was purged by continuous pumping for 10–15 minutes to eliminate stagnant water and obtain representative aquifer conditions. [8], [9]. Borehole depths varied across the basin, reflecting differences in lithological formations and groundwater potential zones.

Samples were filtered through a 0.45 µm membrane filter and preserved with nitric acid immediately after collection, then stored in high-density polyethylene (HDPE) bottles to prevent contamination. Heavy metals (Cd, Cu, Fe, Pb, Mn, Ni, Zn) were analysed using Inductively Coupled Plasma Mass Spectrometry (ICP-MS). Physicochemical parameters (pH, temperature, salinity, dissolved oxygen (DO), total dissolved solids (TDS), electrical conductivity (EC), oxidation-reduction potential (ORP)) were measured in situ using a YSI Pro DSS multi-parameter probe. Major anions (HCO_3, Cl, SO_4, NO_3) and cations (Ca, Mg, K, Na) were determined in the laboratory using titration and ion chromatography methods. To reduce variability, measurements taken following purging were recorded in triplicate and averaged.

Strict QA/QC procedures were implemented throughout sampling and analysis. All instruments, including the YSI Pro DSS multi-probe meter and ICP-MS, were calibrated daily using certified standards. Field blanks, duplicates, and recovery checks were incorporated to monitor precision and accuracy. Exceptionally high Fe and Mn concentrations were subjected to repeat analysis to confirm validity. These values were retained after verification, reflecting true geogenic enrichment rather than analytical error. The consistency of replicate measurements and alignment with known lithological characteristics of the Linggi River Basin further support the authenticity of these extreme values. By explicitly confirming outliers through re-analysis and QA/QC protocols, confidence in the dataset is strengthened and potential bias is minimised [8].

The analytical results were compared against World Health Organisation drinking water standards to evaluate compliance and identify potential risks [9].

2.3. Statistical Analysis

Relationships between water quality parameters and heavy metal concentrations were examined using Pearson's correlation coefficient, a standard approach in groundwater research [10]. This parametric test assesses the strength and direction of linear relationships between continuous variables. Positive correlations may reflect common sources or similar geochemical behaviour, whereas negative correlations suggest differing origins or interactions. For instance, the correlation observed between Pb and Cd ($r = 0.33$, $p < 0.05$) points to shared anthropogenic inputs such as industrial discharge or mining, consistent with findings by Ullah et al. [10].

Correlation analysis is widely applied in hydrochemical studies to identify variable patterns and infer contamination sources. It is particularly effective in distinguishing natural geogenic contributions from human-induced pollution. Microsoft Excel (Data Analysis Toolpak) was used to create descriptive statistics and correlation matrices, while SPSS was

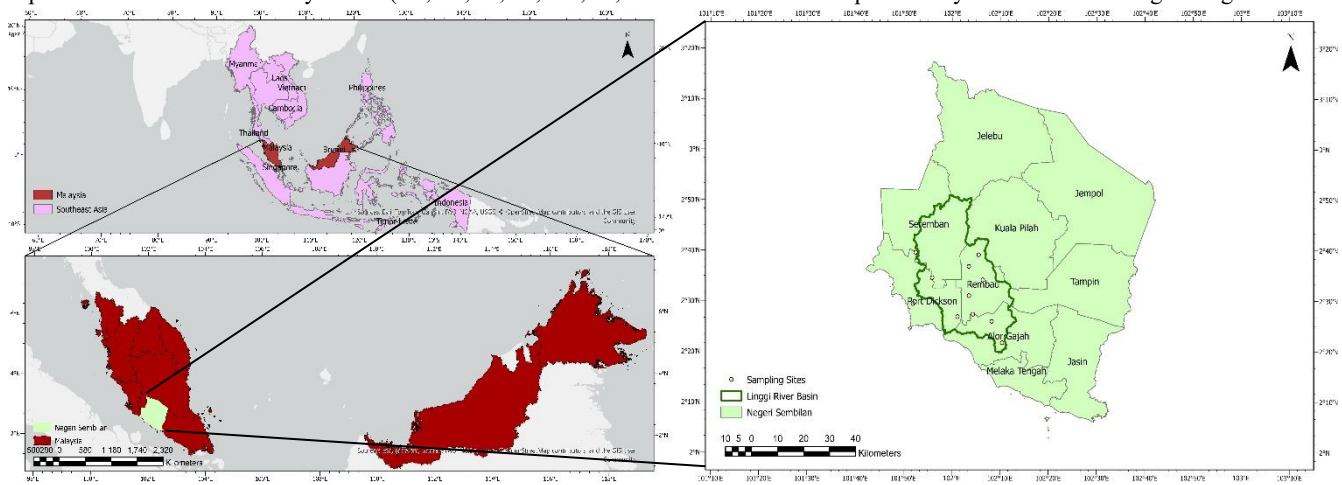

Fig. 1 - Illustrates the basin's location within Southeast Asia, Malaysia and Negeri Sembilan, providing regional context for the study.

used to verify results and guarantee robustness. The effectiveness of Pearson's approach with normally distributed data and a proven track record in environmental monitoring had led to its selection [11].

2.4. Machine Learning

Two predictive models were applied: Random Forest (RF) and Multiple Linear Regression (MLR). RF was chosen for its ability to capture nonlinear relationships and provide high predictive accuracy in environmental datasets[12] At the same time, MLR served as a baseline statistical model [13] The dataset was partitioned into training (80%) and testing (20%) subsets using stratified random sampling to preserve variability across sites. Performance was assessed using three standard regression metrics: coefficient of determination (R^2), root mean squared error (RMSE) and mean absolute error (MAE), which are widely accepted for evaluating predictive accuracy in environmental science[14].

2.4.1. Shapley Additive Explanations (SHAP)

SHAP were used to interpret RF predictions and identify the most influential geochemical drivers of heavy metal mobility. Parameters such as ORP, HCO_3 concentration, and temperature were highlighted as key contributors [14]

2.5. Health Risk Assessment

Health risk assessment (HRA) was conducted separately for adults and children to account for age-specific differences in exposure rates, sensitivity and body weight[15]

2.5.1. Non-Carcinogenic Health Risk Assessment

Heavy metals found in water can be absorbed into the human body primarily through two exposure routes: oral ingestion and dermal absorption. Both pathways can lead to adverse health effects, as outlined in the human health risk assessment framework developed by the United States Environmental Protection Agency [16]. A health risk assessment was conducted to evaluate dermal and oral exposure to contaminated groundwater for both adults and children. The anticipated health risks associated with oral and dermal exposure can be quantified using Equations (1), and (2), and the parameters for these equations are defined in Table 1.

$$CDI_{oral} = \frac{CW \times IR \times EF \times ED}{BW \times AT} \quad (1)$$

$$CDI_{dermal} = \frac{CW \times SA \times PC \times ET \times EF \times ED \times CF}{BW \times AT} \quad (2)$$

Table 1 - Parameters Used to Calculate Chronic Daily Intake (CDI) via Oral and Dermal Pathways in Health Risk Assessment

Parameter	Description	Adults	Children
CDI	Chronic Daily Intake (mg/kg/day)	Calculated	Calculated
CW	Concentration of contaminant in groundwater (mg/L)	Input data	Input data
IR	Ingestion rate (L/day)	2	1
EF	Exposure frequency (days/year)	365	365
ED	Exposure duration (years)	30	6
ET	Exposure time (hours/day)	0.58	1
SA	Skin surface contact area (cm^2)	18,000	6600
CF	Conversion factor (L/cm^3)	0.001	0.001
BW	Body weight (kg)	55	15
AT	Averaging time for non-carcinogenic effects (days)	25550	2190
PC	Dermal Permeability Coefficients (cm/hr)	Refer to Table 2	

Table 2 - Dermal Permeability Coefficients (PC) for Selected Heavy Metals in Groundwater Exposure Assessment

Metal	PC (cm/hr)
Fe	0.001
Pb	0.001
Cd	0.001
Ni	0.001
Mn	0.001
Cu	0.001
Zn	0.0006

The hazard quotient (HQ) for oral and dermal exposure was further calculated using Equation (3):

$$HQ = \frac{Intake}{RfD} \quad (3)$$

Where RfD is the reference dose of specific contaminants (mg/kg/day) obtained from the database of IRIS (Integrated Risk Information System) [16] (Table 3). Further, the total non-carcinogenic health risk, in the form of hazard index (HI), the cumulative sum of HQs for all contaminants, was also estimated. HI ≤ 1 denotes the acceptable non-carcinogenic health risk, and HI > 1 represents that a contaminant may harm human health.

Table 3 - RfD is the reference dose of specific contaminants (mg/kg/day) obtained from the database of IRIS

Metal	Rfd (cm/hr)
Fe	0.7
Pb	0.0035
Cd	0.001
Ni	0.02
Mn	0.7
Cu	0.04
Zn	0.03

2.5.2. Carcinogenic Health Risk Assessment

The probable cancer risks from exposure to a specific contaminant dose in drinking water can be computed using the Excess Cancer Risk (ECR). It is calculated by multiplying the Chronic Daily Intake (CDI) by the Cancer Slope Factor (CSF) as shown in Equation (4) with detailed variable explanations provided in Table 4.

$$ECR = CDI_{oral} \times CSF \quad (4)$$

Table 4 - Parameters for Estimating Excess Cancer Risk (ECR) via Oral Exposure to Heavy Metals in Groundwater

Parameter	Description	Adults	Children
CDI	Chronic Daily Intake (mg/kg/day)	Calculated	Calculated
ECR	Excess cancer risk (unitless)	Calculated	Calculated
CSF	Cancer Slope Factor (mg/kg/day)$^{-1}$	Depend	Depend

Table 5 lists the CSF values applied. Cd and Pb from Salami et al., [15]. Ni lacks an oral/dermal CSF and is excluded unless inhalation exposure is considered. Fe, Mn, Cu and Zn are treated as non-carcinogenic, with CSF not applicable.

Table 5 - CFS (mg/kg/day)$^{-1}$ for each carcinogenic contaminant used in ECR calculation

Heavy Metals (HMs)	CSF (mg/kg/day)$^{-1}$
Cd	6.3
Pb	0.0085

3. Results and Discussion

3.1. Water Quality Characteristics of Groundwater

Groundwater samples from 11 wells across the Linggi River Basin provided insights into physicochemical variability, though the limited sample size constrains spatial representativeness. This restriction increases the risk of overfitting in predictive models, particularly for metals with heterogeneous distributions such as Cd and Ni. Future work should expand sampling coverage, integrate secondary datasets, and apply advanced approaches such as transfer learning or ensemble modelling to improve robustness and generalisability.

Table 6 summarises the parameters' descriptive statistics. In order to evaluate compliance and possible health hazards, the dataset's fifteen physicochemical parameters are compared to WHO drinking water standards [9].

The pH of groundwater varied from 5.04 to 7.74 (mean: 6.70), suggesting conditions that were mildly acidic to almost neutral. Concerns regarding increased solubility and mobility of metals such as Fe, Mn, and Pb were raised by multiple sites (SS7, SS8, SS1, SS10, and SS3) that reported values below the WHO recommendation lower limit of 6.5. Metal dissolution is facilitated by slight acidity, and the lowering of pH may be caused by urbanisation and high human density [17].

The temperature averaged 30.54 °C, which is typical for tropical aquifers [18]. Positive correlations with Fe, Mn, Pb, Zn and Ni ($r = 0.50–0.73$, $p < 0.05$) suggest thermal enhancement of geochemical reactions [19].

Low ionic content was indicated by EC and TDS averages of 235.51 µS/cm and 161.68 mg/L, respectively, which were both below the WHO threshold of 1000 mg/L [6]. Ionic enrichment was validated by the significant correlation between EC and TDS ($r = 0.89$, $p < 0.001$).

Although high readings (up to 120.67 mg/L) may indicate sample artefacts, DO averaged 6.34 mg/L, which is indicative of recent recharge or shallow flow [6]. The ORP showed diverse redox conditions, ranging from –93.91 to +349.00 mV. Positive ORP values stabilise metals, whereas negative values mobilise Fe and Mn [6].

HCO_3^- ranged from 0.81 to 222.85 mg/L (mean: 42.19 mg/L), with SS4 exceeding WHO limits. Cl^- and SO_4^{2-} averaged 18.59 mg/L and 3.98 mg/L, respectively, both below WHO limits, indicating minimal salinity and limited anthropogenic intrusion [20]. Cl^- showed a strong positive correlation with Pb ($r = 0.57$, $p < 0.001$), suggesting possible sewage or urban runoff influence [6] SS4 exceeded WHO standards, and HCO_3- varied from 0.81 to 222.85 mg/L (mean: 42.19 mg/L). The average levels of $Cl-$ and SO_4^{2-} were 18.59 mg/L and 3.98 mg/L, respectively, below WHO guidelines, suggesting low salinity and little anthropogenic intrusion [20]. Cl^- and Pb had a significant positive association ($r = 0.57$, $p < 0.001$), indicating that sewage or urban runoff may have an impact [6]

NO_3 peaked at 41.87 mg/L, which is lower than the 50 mg/L WHO threshold. These high levels could be a sign of sewage intrusion or agricultural runoff, which could cause health problems, including methemoglobinemia in newborns [21].

Na concentrations were much higher than the WHO recommended limit of 50 mg/L (mean: 148.30 mg/L; maximum: 3416.16 mg/L). Elevated Na from fertilisers, wastewater, and geogenic sources may reduce soil permeability and pose health hazards to those with hypertension [8]. K concentrations (mean: 11.59 mg/L; SD: 38.04 mg/L) were significantly lower. Both Mg and Ca showed high standard deviations (SD: 411.33 and 498.01 mg/L, respectively), reflecting regional variability caused by various lithological characteristics [22]. At SS3, Mg concentrations were exceptionally high (2366.86 mg/L), exceeding WHO limits, likely influenced by land use practices [23].

Table 6 - Statistical summary physicochemical

Water Quality Parameter	Unit	Mean	SD	WHO (2022)
pH	-	6.70	0.61	6.5-8.5
Temp	°C	30.54	2.56	Normal ± 2 ◦C
Sal	ppt	0.10	0.07	1000
EC	µS/cm	235.51	152.73	-
DO	ppm	6.34	20.59	-
TDS	mg/L	161.68	113.75	1000
ORP	mV	61.47	100.87	-
HCO_3	mg/L	42.19	62.43	200
Cl	mg/L	18.58	10.98	150
SO_4	mg/L	3.97	5.45	12

NO_3	mg/L	3.60	7.56		200
Ca	mg/L	267.94	498.00	-	
Mg	mg/L	75.63	411.33		500
K	mg/L	11.59	38.04		250
Na	mg/L	148.30	595.06		50

*Temp – temperature, Sal - salinity

Descriptive statistics of all HM pollution records for groundwater across the watershed blocks are presented in Table 7, with spatial aggregation shown in Figure 2. Fe exhibits the highest occurrence in the Linggi River Basin. Followed by Mn, Zn, Cu with Pb and Cd being the least abundant.

Fe concentrations were significantly higher than the WHO aesthetic limit of 0.3 mg/L, with a mean of 164.53 mg/L and a maximum of 1817.76 mg/L. Parallel with observations in the Linggi Basin, elevated Fe indicates significant geogenic input from iron-rich lateritic soils and aquifer-hosted minerals [24], [25].

Concerns on neurotoxicity were raised because Mn concentrations (mean 6.93 mg/L; max 46.49 mg/L) were significantly above the WHO limit of 0.4 mg/L. Negative ORP values encourage the reductive breakdown of Mn oxides, and high Mn levels are associated with decreasing groundwater conditions. There have been reports of similar redox-driven mobilisation in temperate and tropical aquifers [26]. The subsurface environment is rich in both metals (Fe and Mn), with anthropogenic inputs and soil layers next to aquifer clay zones as the main sources. Their higher concentrations in comparison to other HMs are likely driven by this dual origin [27].

While SS9 (max 2.28 mg/l) exceeded the WHO limit, which can have gastrointestinal effects at high levels, Cu (mean 0.21 mg/L) did not. The excess most likely results from pipe deterioration and human inputs from household wastewater [10], [28].

Pb (mean 0.14 mg/L) is higher than the corresponding WHO standard of 0.01 mg/L. The highest concentrations were observed in mixed agricultural-residential zones at SS4 (0.82 mg/L) and SS2 (0.62 mg/L). Potential sources include fertiliser runoff, wastewater discharge, and vehicular emissions. Pb contamination poses significant neurotoxic risks, especially for children [6].

Both Cu and Pb are frequently linked to man-made sources such as industrial discharge, agricultural runoff, and plumbing corrosion. Urbanisation and inadequate waste management techniques, especially leachate migration from unsanitary landfills, have been implicated in causing lead pollution in Malaysia [2].

While Zn was detected in all samples, its concentration remained relatively low (mean: 0.34 mg/L), suggesting minimal anthropogenic influence and likely reflecting natural background levels [29]. Throughout the Linggi River Basin, Ni (mean 0.07 mg/L) and Cd (mean 0.01 mg/L) were identified. Their presence is alarming because of their bioaccumulative nature and propensity to affect aquatic ecosystems and human health adversely, even if their concentrations were generally lower than WHO recommended values (Ni: 0.07 mg/L vs. 0.07 mg/L limit; Cd: 0.01 mg/L vs. 0.003 mg/L limit) [30]. These metals are often associated with industrial activities, electroplating and phosphate fertilizers, which are relevant to the mixed land use observed in the basin (Figure 1).

Elevated Ni values were observed at SS3 (0.65 mg/L) and SS2 (0.22 mg/L), both exceeding the WHO guideline. Sampling site SS3 is positioned near a riverside residential zone and a roadway, near a pumping station; this layout suggests potential contributions from urban runoff and infrastructure corrosion. Meanwhile, SS2 is located near a mosque and agricultural plots, which likely reflects inputs from domestic wastewater and fertiliser application [20].

Although Cd was consistently detected in low levels, there were noticeable abnormalities at SS10 (0.15 mg/L) and SS8 (0.045 mg/L). Phosphate fertilisers and extensive use of agrochemicals are implicated in the land-use context of SS10 (agricultural terrain close to a fishing port. In contrast, the contamination at SS8 is likely attributed to domestic wastewater discharge from the adjacent residential community [31].

Remarkable inter-element correlations between a number of heavy metals were found using correlation analysis. Mineral dissolution processes were indicated by Zn's high positive correlations with Fe ($r = 0.81$, $p < 0.001$), Ni ($r = 0.80$, $p < 0.001$), and Pb ($r = 0.69$, $p < 0.001$) [32]. Cu's moderate associations with Fe ($r = 0.67$, $p < 0.001$), Pb ($r = 0.66$, $p < 0.001$), and Zn ($r = 0.67$, $p < 0.001$) confirmed its association with plumbing corrosion and industrial discharge [2]. The groundwater contamination profile is influenced by a complicated interaction between anthropogenic inputs such as urban runoff, fertiliser use, and waste leachate, and geogenic releases such as mineral weathering and reductive dissolution [10].

Table 7 - Statistical summary of heavy metal

Heavy Metal	Unit	Mean	SD	WHO (2022)
Fe	mg/L	164.53	409.55	0.300
Mn	mg/L	6.93	13.27	0.080
Cu	mg/L	0.21	0.57	2.000
Pb	mg/L	0.14	0.25	0.010
Zn	mg/L	0.34	0.44	3.000-5.000
Ni	mg/L	0.07	0.13	0.070
Cd	mg/L	0.01	0.03	0.003

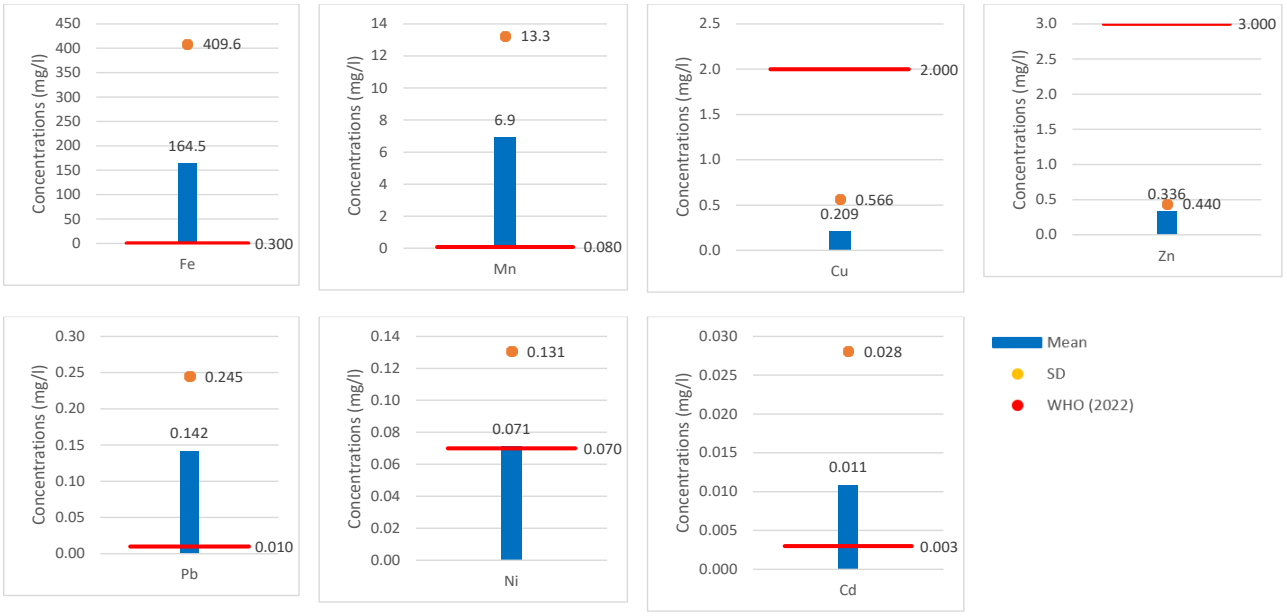

Fig. 2 - Comparison of mean concentrations, standard deviations and WHO (2022) guideline values for selected metals in Linggi River Basin.

3.2. Machine Learning

Table 8 presents a comparative evaluation of RF and MLR models in predicting the concentrations of 7 heavy metals (Fe, Mn, Cu, Pb, Zn, Ni and Cd) based on physicochemical parameters. Model performance was evaluated using R^2, RMSE, and MAE metrics on the test set. In general, the RF model demonstrated greater predicted accuracy and resilience, consistently outperforming MLR across six of the seven heavy metals [32], [33], [34].

For Fe, RF achieved $R^2 = 0.715$ with lower error margins (MAE = 1.247; RMSE = 1.141) compared to MLR (R^2 = 0.489; MAE = 1.670; RMSE = 1.469). RF is especially useful for metals with high variability, particularly Fe, because it can detect nonlinear correlations and limit the influence of outliers [35], [36].

For Mn, MLR performed poorly (R^2 = 0.200; MAE = 1.276; RMSE = 0.854), whereas RF generated a respectable R^2 of 0.471 with lower error metrics (MAE = 1.037; RMSE = 0.780). Both models had comparatively low R2 values for Cu and Pb, but RF still showed superior generality. RF outperformed MLR (Cu = 0.186; Pb = 0.351) with consistently reduced MAE and RMSE values, achieving R^2 scores of 0.342 (Cu) and 0.460 (Pb) [13].

Ni's prediction yielded a more intricate result. RF achieved R^2 = 0.409 with low error margins (MAE = 0.065; RMSE = 0.054), while MLR failed with a highly negative R^2 (−3.441). This implies that while MLR's linear assumptions restrict performance, Ni concentrations show moderate variability and nonlinear interactions that RF can partly address [13].

Remarkably, Zn was the only metal where MLR performed relatively better than RF in terms of R^2 (MLR = 0.853 vs. RF = 0.814), despite the fact that RF maintained comparable error metrics (MAE = 0.140; RMSE = 0.102). This implies that Zn concentrations could demonstrate more linear patterns, which makes MLR a realistic alternative for this particular metal [37], [38].

Both models performed poorly for Cd, with RF yielding $R^2 = -47.372$ and MLR with $R^2 = -22.422$. Despite low MAE and RMSE values, these results are misleading, as both models failed to explain variance. Cd prediction challenges likely stem from (i) imbalanced distribution across sites, (ii) low variance reducing signal strength, and (iii) weak correlation between Cd and measured physicochemical predictors [37], [38].

Generally, these findings highlight how crucial model selection is for environmental prediction tasks. RF is an effective technique for environmental monitoring and geographic risk assessment because of its capacity to capture nonlinear connections and generalise across a variety of metals, particularly when working with complex and heterogeneous datasets. Similar findings were noted in Liang et al.'s ensemble learning studies [30], where RF continuously produced better accuracy and resilience while modelling nonlinear contamination patterns [37], [38]

To mitigate these limitations, future studies should expand the predictor set to include land-use variables, soil composition, and hydrogeological factors that more directly influence Cd mobility. Increasing the number of sampling sites would also improve statistical robustness and reduce overfitting. Advanced modelling approaches such as ensemble learning, feature engineering, or hybrid models integrating geospatial data may further enhance predictive accuracy. While RF demonstrated superior performance for metals with higher variability. Cd remains challenging to model due to its site-specific drivers and limited dataset representation. These findings highlight the importance of careful predictor selection and the need for broader datasets when modelling trace contaminants with heterogeneous sources.

Predicted versus observed line plots, which are frequently used in environmental modelling to evaluate alignment and identify systemic bias, were created to illustrate model performance [39]. Furthermore, the RF model was subjected to SHAP (SHapley Additive exPlanations) analysis in order to evaluate feature contributions [40], [41].

Table 8 - Performance Comparison of Random Forest and Multiple Linear Regression Models for Heavy Metal Prediction.

HMs	R^2 Score		MAE		RMSE	
	RF	MLR	RF	MLR	RF	MLR
Fe	0.715	0.489	1.247	1.670	1.141	1.469
Mn	0.471	0.200	1.037	1.276	0.780	0.854
Cu	0.342	0.186	0.328	0.364	0.180	0.237
Pb	0.460	0.351	0.152	0.166	0.076	0.099
Zn	0.814	0.853	0.140	0.124	0.102	0.113
Ni	0.409	-3.441	0.065	0.177	0.054	0.101
Cd	-47.372	-22.422	0.027	0.019	0.016	0.015

3.2.1. Shapley Additive Explanations

The SHAP summary plot provides a powerful lens through which to interpret the behaviour of machine learning models, particularly in environmental data analysis [14]. SHAP analysis was applied to the RF models to interpret feature contributions and identify geochemical drivers of heavy metal mobility. By measuring the impact of various variables on predictions, this method improves transparency and converts complex model behaviour into useful information for public health and groundwater quality evaluation [42].

The horizontal location of each dot in the SHAP plot indicates the strength and direction of the feature's influence on the prediction; each dot represents a SHAP value for a particular instance. The colour gradient illustrates how features impact results, going from blue (low feature value) to red (high feature value) (Figure 3).

Based on SHAP analysis, three features (ORP, HCO_3^-, and temperature) consistently emerged as influential across all seven RF models created for heavy metal prediction and consistently consistently emerged as the most influential predictors across all models. Aquifer redox status is controlled by ORP, which stabilises Cu and Pb under oxidising conditions while mobilising Fe, Mn, Ni, and Cd under reducing conditions [43], [44]. HCO_3^- facilitates metal transport and carbonate complexation by acting as a buffering agent [45], [46]. Temperature accelerates geochemical processes that free metals from sediments by modifying microbial activity and reaction kinetics [47].

With the exception of Fe, Cl^- was maintained as a predictor in every model, indicating its broad significance in affecting metal mobility. It is known that Cl^- can form soluble complexes with metals such as Pb, Cd, and Cu, especially in saline or human-influenced environments [47].

In the Fe, Zn, and Ni models, Mg and Na emerged as predictors. Mg affects sorption competition and ionic strength, whereas Na^+ contextualises ionic environments by acting as a stand-in for anthropogenic inputs and salinity [48], [49].

In the Mn, Zn, and Ni models, TDS emerged as a potent predictor. TDS, which refers to the total concentration of dissolved ions in groundwater, is frequently used as a stand-in for the degree of pollution or the intensity of the water-rock interaction. Its appearance in these three models indicates that the mobilisation of Mn, Zn, and Ni is highly sensitive to solute load and cumulative ionic strength [12].

The Mn, Cu, and Zn models maintained the characteristic pH, indicating their crucial function in regulating metal solubility. By encouraging their disintegration and decreasing adsorption to mineral surfaces, acidic environments typically promote the mobility of metals like Cu and Zn. According to experimental research, Zn and Cu are more soluble at lower pH levels because of less surface complexation and improved desorption from mineral [50].

In line with its function in co-dissolution and buffering, Ca was preserved in the Mn and Zn models. In weathering zones, calcium is frequently liberated from carbonate minerals and frequently coexists with trace metals [14].

K only showed up in the Ni model, suggesting a narrow yet distinct function. K may co-occur with Ni in mineral dissolution zones or indicate anthropogenic inputs such as fertiliser runoff, even though it does not strongly interact with the majority of heavy metals [51].

Only the Mn model retained EC. EC is frequently employed as a rapid indicator of contamination or mineralisation since it represents the overall ionic concentration of groundwater. Because manganese is redox-active and has a propensity to adsorb or precipitate under different conductivity regimes, its appearance in the Mn model indicates that manganese mobility is especially susceptible to changes in ionic strength [52].

Overall, the results show that SHAP analysis effectively reveals both direct and indirect geochemical controls on metal behaviour. The selective emergence of features across models reflects the unique speciation, transport mechanisms and environmental sensitivities of each metal. Moreover, the interplay between ionic strength indicators (TDS, EC, Na^+), buffering agents (HCO_3, Ca) and redox controls (ORP) underscores the multidimensional nature of groundwater geochemistry and its influence on predictive modelling.

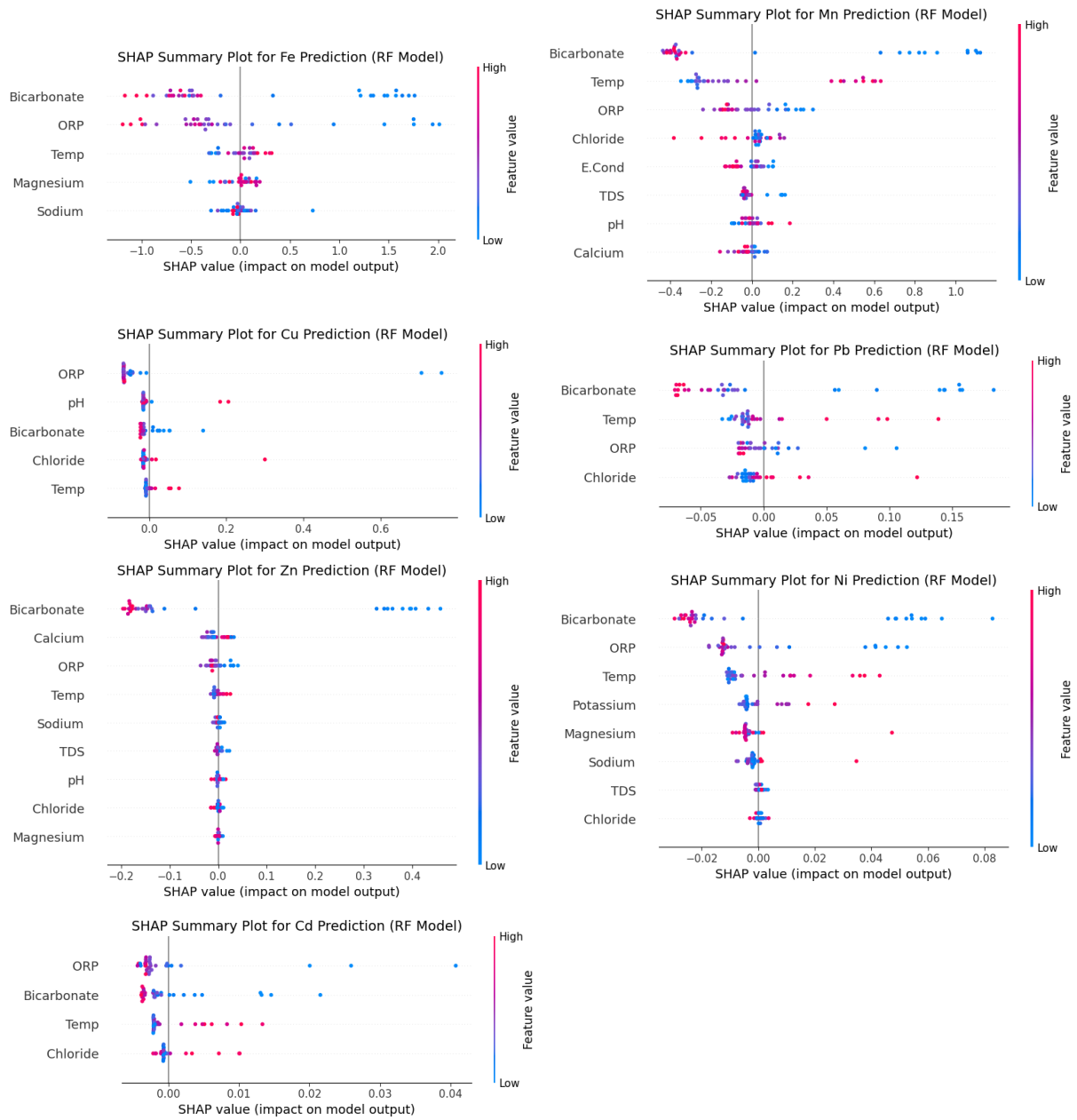

Fig. 3 - Shapley additive explanations (SHAP) values in predicting the risks for all the input descriptors, with all the data points included, reveal the most influential features for(a) Fe, (b) Mn, (c) Cu, (d) Pb, (e) Zn, (f) Ni and (g) Cd.

Health Risk Assessment

3.2.2. Non-Carcinogenic Risk Assessment

The non-carcinogenic health risks associated with shallow groundwater exposure in the Linggi River Basin were assessed via oral ingestion and dermal contact pathways, targeting both adult and child populations across residential, agricultural and mixed-use zones.

Dermal HI values remained well below the threshold of 1.0, with mean values of 0.0254 for adults and 0.1372 for children, confirming minimal risk from bathing or skin contact. Even the highest child dermal HI values (SS9, SS7, SS3) were under 0.25 [15].

Oral ingestion posed far greater concern, especially for children, whose HI values were 2–4 times higher than those of adults and frequently exceeded 1.0. The highest HI was at SS9 has the highest HI (64.15), followed by SS7, SS3, SS4, SS2, SS8, SS10, SS5, SS11, SS1, and SS6. Fe, Pb, Cd, Cu, and Zn were the primary causes of elevated indices, underscoring the necessity of child-focused interventions in high-risk areas [53].

In 64% of sites, Hazard Quotient (HQ) Fe is dominant, accounting for more than half of the cumulative risk in SS1, SS2, SS3, SS4, SS7, SS9, and SS11. Geogenic enrichment and human inputs like damaged pipelines and fertiliser runoff are examples of sources [53].

Pb HQ is significant in 55% of sites (SS2, SS4, SS6, SS7, SS8, SS9), with HQ values exceeding 1. Residential contamination stems from waste disposal and septic leakage, while agricultural zones reflect pesticide and fertiliser use. Pb's neurotoxicity makes child exposure particularly concerning [54], [55].

Mn HQ is elevated in 36% of sites (SS1, SS2, SS4, SS5). Though concentrations were lower than Fe and Pb, co-

occurrence amplified cumulative HI. Sources include wastewater leakage and fertiliser mobilisation [54], [55].

Due to their lower body mass, higher water intake per kilogram, and longer exposure times, children consistently displayed higher HI values. This pattern mirrors findings from other regions (Raheja et al. [55], Mohammed et al. [54], Xie et al. [56], Salami et al. [15]).

The highest oral HI values were found in agricultural zones (SS4, SS10, and SS11), which are associated with animal runoff and fertiliser. Domestic and agricultural activities had cumulative effects in mixed-use areas (SS7). There were isolated Pb and Cd surges at residential sites close to septic tanks or waste disposal (SS9, SS8). Oral HI maps highlight SS9, SS7, and SS3 as hotspots for child risk.

While dermal exposure poses minimal risk, oral ingestion particularly for children represents a significant non-carcinogenic hazard. To protect vulnerable populations and guarantee sustainable groundwater use, targeted interventions including buffer enforcement, seasonal monitoring, household water treatment, and public education are crucial [22].

3.2.3. Carcinogenic Risk Assessment

Excess Cancer Risk (ECR) values for oral exposure to Pb and Cd, two heavy metals that can cause cancer if consumed for an extended period of time. The USEPA threshold of 1x10-4, which establishes the acceptable lifetime cancer risk level, was used as a guideline for these values. Cd continuously showed higher ECR values than Pb across the Linggi River Basin, frequently over the threshold, especially in agricultural zones. The ECR values, which are used to assess the possible health concerns associated with heavy metal exposure, are also statistically summarised in the table.

Because of their lower body weight, higher water intake per unit mass, and longer exposure times, children had significantly higher ECR values than adults at every sampling location (Figure 5). SS10 (Agriculture) had the highest ECR values, with adults reaching 5.12×10^{-3} and children reaching 2.19×10^{-2}, well above the allowable limit. Children's ECR values in SS8 and SS2 (Residential) exceeded 6.68×10^{-3} and 6.43×10^{-3}, respectively, indicating increased dangers.

Over 90% of the cumulative risk was accounted for by Cd, which dominated the overall ECR across all sites. Pb, on the other hand, contributed very little but exceeded permissible limits in a number of places. This pattern is consistent with research by Salami et al. [57], Zhu et al. [58], and Kumar et al. [22], who found that Cd and Pb were the main carcinogens in groundwater in China and Indonesia [15], [22], [58]. Their research highlighted Cd's anthropogenic origins

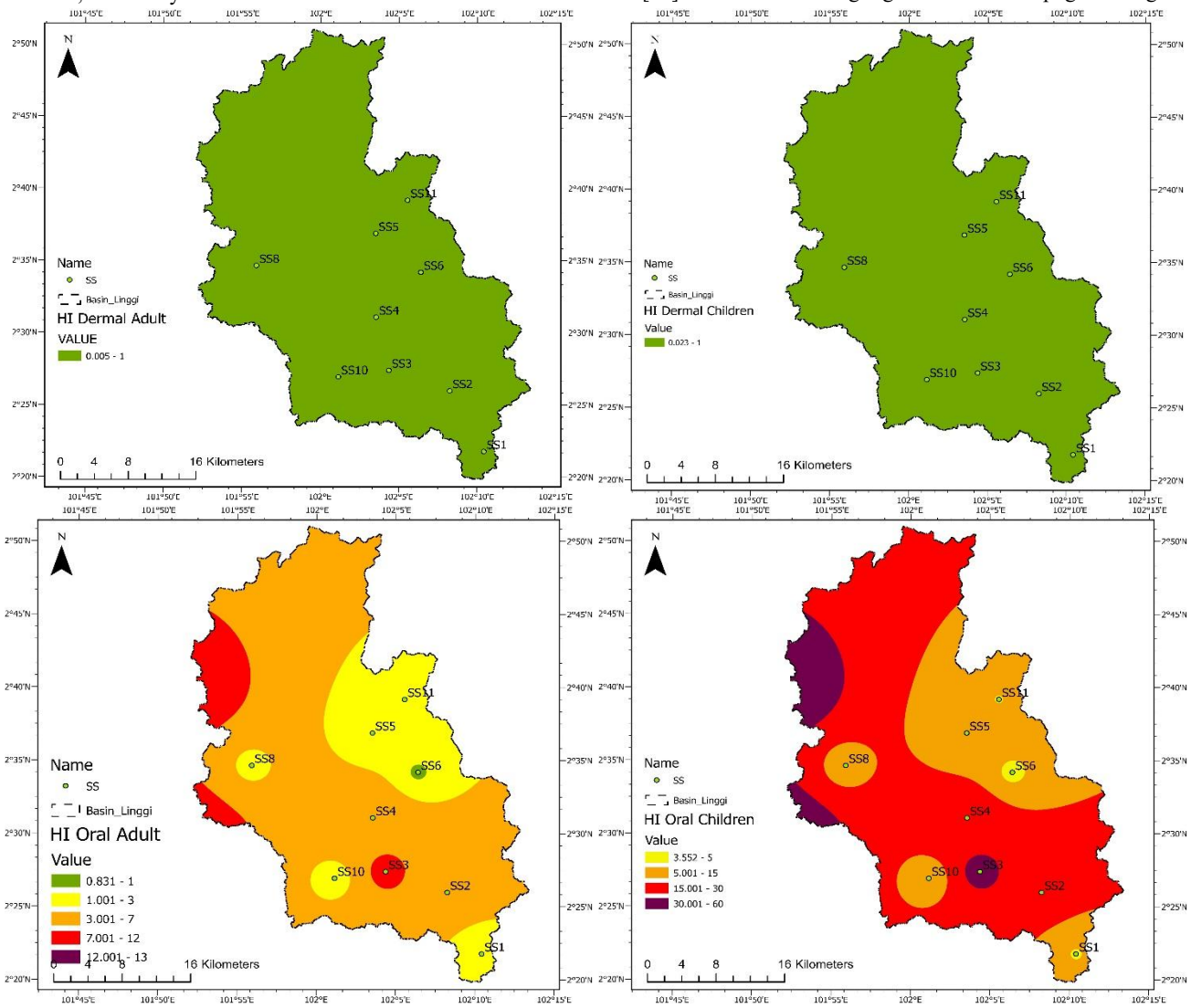

Fig. 4 - The spatial distribution of Hazard Index (HI) values for oral and dermal exposure pathways across adult and child populations in the Linggi River Basin.

Published by JPS Publishing
https://journal.water.gov.my

and geochemical persistence, especially from phosphate fertilisers, industrial discharge, and urban runoff, underscoring its importance as a public health concern. Salami et al. [57] also reported elevated ECR values for Cd and Cr in shallow groundwater near the Citarum River, with children showing significantly higher risks due to physiological sensitivity and exposure patterns [15].

Spatially, agricultural sites (SS4, SS10, SS11) exhibited the highest cumulative ECR values, likely due to the use of phosphate fertilisers and agrochemicals containing trace amounts of Cd. This pattern is consistent with findings from Raheja et al. [55] and Mohammed et al. [95], who reported elevated carcinogenic risks in agricultural zones linked to fertiliser runoff and shallow well vulnerability [54], [55]. Residential zones near urban centres (SS2, SS8, SS9) also showed elevated risks, echoing observations of contamination from domestic waste, septic tank leakage, and aging infrastructure.

The findings underscore the urgent need for groundwater quality management, particularly in high-risk zones. This includes promoting alternative water sources, implementing household treatment technologies, and enhancing public education to mitigate long-term exposure. Elevated ECR values for children in SS10, SS8, and SS2 reinforce the need for stricter regulation of agricultural inputs, seasonal monitoring, and child-focused health interventions to safeguard vulnerable populations.

with SHAP analysis identifying ORP, HCO_3, and temperature as dominant geochemical drivers. Health risk assessment confirmed oral ingestion as the primary exposure pathway, with children more vulnerable than adults. More than 90% of the cumulative Excess Cancer Risk was attributed to cadmium, which was the primary cause of carcinogenic risk, especially in agricultural and mixed-use areas (Figure 6).

These findings underscore the urgent need for targeted monitoring, stricter land-use management, and improved wastewater treatment to safeguard groundwater-dependent communities. Integrating machine learning with conventional hydrochemical analysis provides a powerful tool for predictive risk management and sustainable groundwater governance. Future research should expand sampling campaigns across monsoon seasons and include emerging contaminants such as pesticides and pharmaceuticals to provide a more comprehensive risk profile.

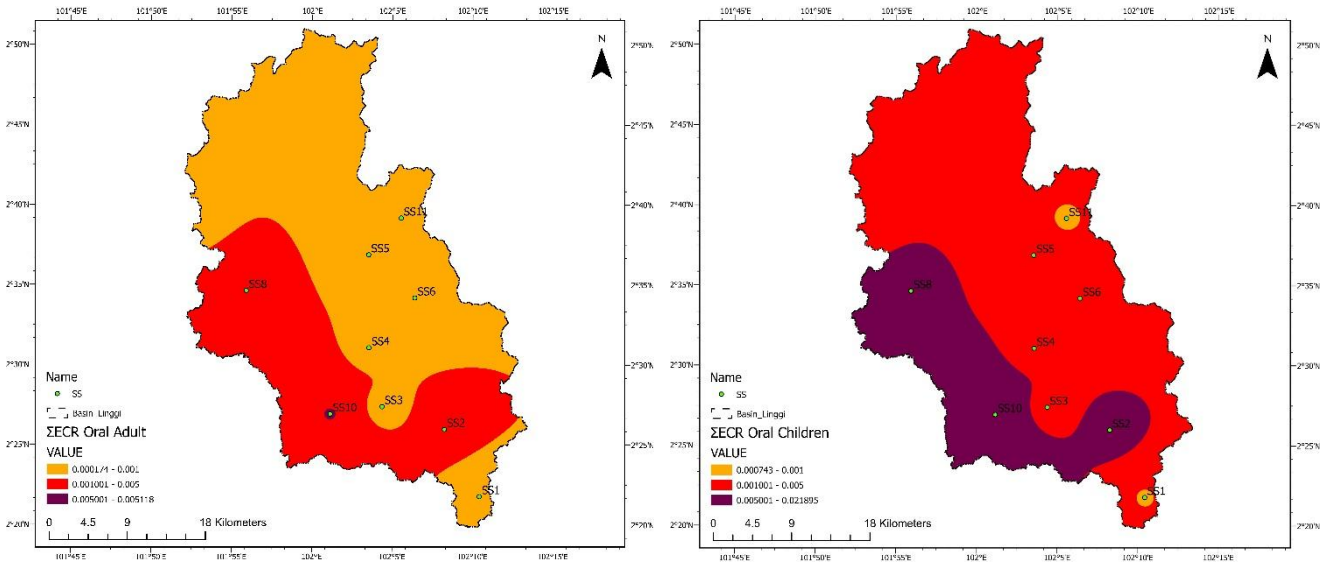

Fig. 5 - Spatial distribution of Excess Cancer Risk (ECR) values for oral exposure to heavy metals in adults and children across the Linggi River Basin.

4. Conclusion

This study evaluated groundwater quality in the Linggi River Basin, applied machine learning models for predictive analysis, and assessed associated health risks. Physicochemical assessment revealed that while most parameters were within WHO standards, several sites showed deviations, particularly in pH, Na, and redox conditions, which enhanced the mobility of Fe, Mn, and Pb. Heavy metals such as Fe and Mn consistently exceeded guideline values, while Pb and Cd contamination posed serious neurotoxic and carcinogenic risks.

Random Forest models outperformed Multiple Linear Regression in predicting nonlinear interactions among metals,

5. Online License Transfer

Not applicable at the moment

Acknowledgement

This study was supported by Universiti Putra Malaysia, specifically through the GP-IPS/2023/9768700, which was instrumental in funding it.

References

[1] R. Das, N. Subba Rao, H. K. Sahoo, and G. Sakram, "Nitrate contamination in groundwater and its health implications in a semi-urban region of Titrol block, Jagatsinghpur district, Odisha, India," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 132, p. 103424, Dec. 2023, doi: 10.1016/j.pce.2023.103424.

[2] W. L. Wong, M. Emre, and G. Talukdar, "Groundwater Contamination by Heavy Metals in Malaysia: Sources, Transport, and Remediation Strategies," *Tropical Environment, Biology, and Technology*, vol. 2, no. 2, pp. 93–105, Oct. 2024, doi: 10.53623/TEBT.V2I2.492.

[3] A. Ismanto *et al.*, "Groundwater contamination status in Malaysia: level of heavy metal, source, health impact, and remediation technologies," *Bioprocess and Biosystems Engineering 2022 46:3*, vol. 46, no. 3, pp. 467–482, Dec. 2022, doi: 10.1007/S00449-022-02826-5.

[4] J. Meng *et al.*, "A framework for risk assessment of groundwater contamination integrating hydrochemical, hydrogeological, and electrical resistivity tomography method," *Environmental Science and Pollution Research 2024 31:19*, vol. 31, no. 19, pp. 28105–28123, Mar. 2024, doi: 10.1007/s11356-024-33030-5.

[5] "Laman Web Rasmi | NAHRIM." Accessed: Feb. 27, 2026. [Online]. Available: https://www.nahrim.gov.my/

[6] A. M. Zainudin *et al.*, "Groundwater potability assessment through integration of pollution index of groundwater (PIG) and groundwater quality index (GWQI) in Linggi River Basin, Negeri Sembilan, Malaysia," *Groundw. Sustain. Dev.*, vol. 26, p. 101225, Aug. 2024, doi: 10.1016/j.gsd.2024.101225.

[7] "Cadmium | UNEP - UN Environment Programme." Accessed: Feb. 27, 2026. [Online]. Available: https://www.unep.org/topics/chemicals-and-pollution-action/pollution-and-health/heavy-metals/cadmium?utm_source=copilot.com

[8] "Procedures for Groundwater Sampling in the Laboratory Services and Applied Science Division | US EPA." Accessed: Feb. 27, 2026. [Online]. Available: https://www.epa.gov/quality/procedures-groundwater-sampling-laboratory-services-and-applied-science-division

[9] "Guidelines for drinking-water quality: fourth edition incorporating the first and second addenda." Accessed: Feb. 27, 2026. [Online]. Available: https://www.who.int/publications/i/item/9789240045064?utm_source=copilot.com

[10] Z. Ullah *et al.*, "Groundwater contamination through potentially harmful metals and its implications in groundwater management," *Front. Environ. Sci.*, vol. 10, p. 1021596, Nov. 2022, doi: 10.3389/FENVS.2022.1021596/BIBTEX.

[11] K. Naresh Dwivedi and A. Lakhera, "A REVIEW OF VARIOUS METHODS OF CORRELATIONS ANALYSIS ON WATER QUALITY PARAMETERS," *International Journal of Research Publication and Reviews Journal homepage: www.ijrpr.com*, vol. 3, no. 5, pp. 1613–1616, 2022, Accessed: Dec. 01, 2025. [Online]. Available: www.ijrpr.com

[12] L. Xu, X. Cui, J. Bian, Y. Wang, and J. Wu, "Dynamic change and driving response of shallow groundwater level based on random forest in southwest Songnen Plain," *J. Hydrol. Reg. Stud.*, vol. 53, no. 1, p. 101800, Jun. 2024, doi: 10.1016/j.ejrh.2024.101800.

[13] P. Dargahi, S. Nasseri, M. Hadi, R. N. Nodehi, and A. H. Mahvi, "Prediction models for groundwater quality parameters using a multiple linear regression (MLR): a case study of Kermanshah, Iran," *Journal of Environmental Health Science and Engineering 2022 21:1*, vol. 21, no. 1, pp. 63–71, Dec. 2022, doi: 10.1007/s40201-022-00836-9.

[14] L. Wang *et al.*, "Machine learning-based analysis of heavy metal contamination in Chinese lake basin sediments: Assessing influencing factors and policy implications," *Ecotoxicol. Environ. Saf.*, vol. 283, no. 9, p. 116815, Sep. 2024, doi: 10.1016/j.ecoenv.2024.116815.

[15] I. R. S. Salami, N. A. Thufailah, N. Fahimah, and D. Roosmini, "Health risk assessment of physicochemical and heavy metals exposures of the usage of shallow groundwater located at the proximity to Citarum River, Indonesia," *Case Studies in Chemical and Environmental Engineering*, vol. 11, p. 101153, Jun. 2025, doi: 10.1016/J.CSCEE.2025.101153.

[16] I. Final, "Risk Assessment: Guidance for Superfund Volume 1 Human Health Evaluation Manual (Part A)," 1989, Accessed: Feb. 27, 2026. [Online]. Available: http://www.epa.gov/swerrims/riskassessment/risk_superfund.html

[17] D. A. Ayejoto and J. C. Egbueri, "Human health risk assessment of nitrate and heavy metals in urban groundwater in Southeast Nigeria," *Ecological Frontiers*, vol. 44, no. 1, pp. 60–72, Feb. 2024, doi: 10.1016/J.CHNAES.2023.06.008.

[18] H. Neidhardt and W. Shao, "Impact of climate change-induced warming on groundwater temperatures and quality," *Applied Water Science 2023 13:12*, vol. 13, no. 12, pp. 235-, Nov. 2023, doi: 10.1007/s13201-023-02039-5.

[19] M. O. Eyankware, O. C. Akakuru, W. O. Osisanya, S. O. Umayah, and K. P. Ukor, "Assessment of heavy metal pollution on groundwater quality in the Niger Delta Region of Nigeria," *Sustain. Water Resour. Manag.*, vol. 9, no. 6, pp. 1–20, Dec. 2023, doi: 10.1007/S40899-023-00955-7/FIGURES/5.

[20] A. J. Tesoriero, S. A. Wherry, D. I. Dupuy, and T. D. Johnson, "Predicting Redox Conditions in Groundwater at a National Scale Using Random Forest Classification," *Environ. Sci. Technol.*, vol. 58, no. 11, pp. 5079–5092, Mar. 2024, doi: 10.1021/acs.est.3c07576.

[21] H. I. Shuval and N. Gruener, "INFANT METHEMOGLOBINEMIA AND OTHER HEALTH EFFECTS OF NITRATES IN DRINKING WATER," pp. 183–193, Jan. 2013, doi: 10.1016/B978-1-4832-1344-6.50017-4.

[22] P. Kumar, P. K. Khan, and A. Kumar, "Health risk assessment upon exposure to groundwater arsenic among individuals of different sex and age groups of Vaishali district, Bihar (India)," *Toxicol. Rep.*, vol. 14, p. 102024, Jun. 2025, doi: 10.1016/J.TOXREP.2025.102024.

[23] G. P.G.M.S., V. P. Sangasumana, and Patabandi. K.P.L.N, "Evaluation of Groundwater Quality and Potability Based on Land Use Variations: A Case Study from Mahiyanganaya, Sri Lanka," *International Journal of Research and Innovation in Applied Science*, vol. X, no. VI, pp. 1422–1436, 2025, doi: 10.51584/ijrias.2025.100600109.

[24] M. D. Suhaimi Bin and E. Phd, "UNIVERSITI TEKNOLOGI MARA DISTRIBUTION AND SOURCE IDENTIFICATION OF HEAVY METALS, TRACE AND RARE EARTH ELEMENTS IN SURFACE SEDIMENT OF LINGGI RIVER," 2021.

[25] K. Krishnan, Y. C. Kong, S. Suganthi, and P. Balu, "Evaluation of heavy metal contamination in sediments of the Linggi River, Malaysia,"

International Journal of Innovative Research and Scientific Studies, vol. 8, no. 11, pp. 196–203, Nov. 2025, doi: 10.53894/ijirss.v8i11.10834.

[26] Z. Zhang, C. Xiao, O. Adeyeye, W. Yang, and X. Liang, "Source and Mobilization Mechanism of Iron, Manganese and Arsenic in Groundwater of Shuangliao City, Northeast China," *Water 2020, Vol. 12, Page 534*, vol. 12, no. 2, p. 534, Feb. 2020, doi: 10.3390/W12020534.

[27] M. S. Islam and M. G. Mostafa, "Occurrence, Source, and Mobilization of Iron, Manganese, and Arsenic Pollution in Shallow Aquifer," *Geofluids*, vol. 2023, no. 1, p. 6628095, Jan. 2023, doi: 10.1155/2023/6628095.

[28] I. I. Ruzi *et al.*, "Heavy metal contamination in Sungai Petani, Malaysia: a wastewater-based epidemiology study," *J. Water Health*, vol. 22, no. 6, pp. 953–966, Jun. 2024, doi: 10.2166/wh.2024.241.

[29] A. M. Zainudin *et al.*, "Groundwater potability assessment through integration of pollution index of groundwater (PIG) and groundwater quality index (GWQI) in Linggi River Basin, Negeri Sembilan, Malaysia," *Groundw. Sustain. Dev.*, vol. 26, p. 101225, Aug. 2024, doi: 10.1016/J.GSD.2024.101225.

[30] C. Liang *et al.*, "Prediction of 3D spatial distribution, driving factors, and risk for heavy metal(loid)s in sediment of Dongdagou River based on ensemble learning," *J. Hazard. Mater.*, vol. 496, p. 139500, Sep. 2025, doi: 10.1016/j.jhazmat.2025.139500.

[31] M. S. Achary *et al.*, "Accumulation and risk assessment of heavy metals employing species sensitivity distributions in Linggi River, Negeri Sembilan, Malaysia," *Ecotoxicol. Environ. Saf.*, vol. 211, p. 111905, Mar. 2021, doi: 10.1016/j.foodcont.2016.04.028.

[32] J. Kim, Y. Ma, and R. M. Maxwell, "Integrating groundwater pumping data with regression-enhanced random forest models to improve groundwater monitoring and management in a coastal region," *Frontiers in Water*, vol. 6, p. 1509945, Dec. 2024, doi: 10.3389/frwa.2024.1509945.

[33] A. Madani, · Mohammed Hagage, · Salwa, and F. Elbeih, "Random Forest and Logistic Regression algorithms for prediction of groundwater contamination using ammonia concentration," *Arabian Journal of Geosciences 2022 15:20*, vol. 15, no. 20, pp. 1619-, Oct. 2022, doi: 10.1007/s12517-022-10872-2.

[34] M. Fiedler, "Using Random Forest Regression to Model the Spatial Distribution of Concentrations of Selected Metals in Groundwater in Forested Areas of the Wielkopolska National Park, Poland," *Forests 2024, Vol. 15,* vol. 15, no. 12, Dec. 2024, doi: 10.3390/f15122191.

[35] C. Singla, R. Aggarwal, S. Kaur, and R. Sharma, "Artificial intelligence-based approach to study the impact of climate change and human interventions on groundwater fluctuations," *AQUA - Water Infrastructure, Ecosystems and Society*, vol. 72, no. 8, pp. 1404–1414, Aug. 2023, doi: 10.2166/AQUA.2023.009.

[36] M. A. Saleh and H. M. Rasel, "Machine learning for groundwater levels: uncovering the best predictors," *Sustainable Water Resources Management 2024 10:5*, vol. 10, no. 5, pp. 166-, Aug. 2024, doi: 10.1007/s40899-024-01146-8.

[37] R. Proshad, M. G. Kibria, D. Khurram, A. Takai, T. Katsumi, and M. A. Abedin, "Modeling heavy metal contamination in fire-affected soils using machine learning in the Kutupalong Rohingya camp, Bangladesh," *J. Environ. Chem. Eng.*, vol. 13, no. 5, p. 117975, Oct. 2025, doi: 10.1016/J.JECE.2025.117975.

[38] H. Hu *et al.*, "Machine learning combined with geodetector to predict the spatial distribution of soil heavy metals in mining areas," *Science of The Total Environment*, vol. 959, p. 178281, Jan. 2025, doi: 10.1016/J.SCITOTENV.2024.178281.

[39] Y. Wang, X. Wu, and M. Hao, "Hyperspectral Inversion and Analysis of Soil Heavy Metals Based on Optimizing CARS Combined with Random Forest," *AGUFM*, 2023, Accessed: Mar. 01, 2026. [Online]. Available: https://scixplorer.org/abs/2023AGUFMGH02...06W/abstract

[40] Z. He and W. Yuan, "Exploring the Influencing Factors of Net Ecosystem Productivity (NEP) Based on Random Forest and SHAP," *Academic Journal of Science and Technology*, vol. 12, no. 2, pp. 242–248, Sep. 2024, doi: 10.54097/404zt095.

[41] H. Gholami *et al.*, "An interpretable (explainable) model based on machine learning and SHAP interpretation technique for mapping wind erosion hazard," *Environmental Science and Pollution Research*, vol. 31, no. 56, pp. 64628–64643, Dec. 2024, doi: 10.1007/S11356-024-35521-X/FIGURES/9.

[42] E. Scornet and G. Hooker, "Theory of Random Forests: A Review", Accessed: Oct. 16, 2025. [Online]. Available: https://hal.science/hal-05006431v1

[43] N. N. Thanh, S. Chotpantarat, N. T. Ha, and N. H. Trung, "Determination of conditioning factors for mapping nickel contamination susceptibility in groundwater in Kanchanaburi Province, Thailand, using random forest and maximum entropy," *Environ. Geochem. Health*, vol. 45, no. 7, pp. 4583–4602, Jul. 2023, doi: 10.1007/S10653-023-01512-Z/FIGURES/9.

[44] A. Kubier, K. Hamer, and T. Pichler, "Cadmium Background Levels in Groundwater in an Area Dominated by Agriculture," *Integr. Environ. Assess. Manag.*, vol. 16, no. 1, pp. 103–113, Jan. 2020, doi: 10.1002/IEAM.4198.

[45] P. I. Chou, X. Wu, Z. Gao, Y. Zhu, and Y. S. Jun, "Bicarbonate concentrations affect arsenic release from arsenopyrite and nanoscale iron(III) (hydr)oxide formation: importance of unconfined aquifer carbonate chemistry," *Environ. Sci. Nano*, vol. 12, no. 6, pp. 3047–3060, Jun. 2025, doi: 10.1039/d4en00805g.

[46] D. Li *et al.*, "Influence of bicarbonate on the abundance of microbial communities capable of reducing U(VI) in groundwater," *RSC Adv.*, vol. 7, no. 78, pp. 49745–49752, Oct. 2017, doi: 10.1039/c7ra09795f.

[47] P. ; Zhang *et al.*, "Water Quality Degradation Due to Heavy Metal Contamination: Health Impacts and Eco-Friendly Approaches for Heavy Metal Remediation," *Toxics 2023, Vol. 11, Page 828*, vol. 11, no. 10, p. 828, Sep. 2023, doi: 10.3390/TOXICS11100828.

[48] "January-February 2023 Groundwater highlights." Accessed: Feb. 28, 2026. [Online]. Available: https://www.ngwa.org/publications-and-news/january-february-2023-groundwater-highlights?utm_source=copilot.com

[49] M. Macena, H. Pereira, L. Cruz-Lopes, L. Grosche, and B. Esteves, "Competitive Adsorption of Metal Ions by Lignocellulosic Materials: A Review of Applications, Mechanisms and Influencing Factors," *Separations 2025, Vol. 12, Page 70*, vol. 12, no. 3, p. 70, Mar. 2025, doi: 10.3390/SEPARATIONS12030070.

[50] S. Zhao, Y. Zhao, Z. Cui, H. Zhang, and J. Zhang, "Effect of pH, Temperature, and Salinity Levels on Heavy Metal Fraction in Lake Sediments," *Toxics 2024, Vol. 12, Page 494*, vol. 12, no. 7, p. 494, Jul. 2024, doi: 10.3390/TOXICS12070494.

[51] A. Davraz and F. Aksever, "Geochemical processes in groundwater chemistry: a case study," *Carbonates and Evaporites 2025 40:3*, vol. 40, no. 3, pp. 92-, Jun. 2025, doi: 10.1007/s13146-025-01124-x.

[52] M. S. Islam and M. G. Mostafa, "Occurrence, Source, and Mobilization of Iron, Manganese, and Arsenic Pollution in Shallow Aquifer," *Geofluids*, vol. 2023, no. 1, p. 6628095, Jan. 2023, doi: 10.1155/2023/6628095.

[53] O. Urme, A. S. Reza, M. I. Adham, and G. S. Sattar, "Arsenic, manganese, and iron concentration in groundwater of northwestern part of Bangladesh using self-organizing maps: Implication for health risk assessment," *Heliyon*, vol. 11, no. 2, p. e41805, Jan. 2025, doi: 10.1016/J.HELIYON.2025.E41805.

[54] M. A. A. Mohammed, N. P. Szabó, E. D. Suliman, M. M. S. Siddig, M. N. M. Hassan, and P. Szűcs, "Integrated assessment of human health risks from groundwater pollutants in Nubian Aquifer, Sudan: Combining source apportionment and probabilistic analysis," *Environmental Challenges*, vol. 20, p. 101176, Sep. 2025, doi: 10.1016/J.ENVC.2025.101176.

[55] H. Raheja, A. Goel, and M. Pal, "Assessment of groundwater quality and human health risk from nitrate contamination using a multivariate statistical analysis," *J. Water Health*, vol. 22, no. 2, pp. 350–366, Feb. 2024, doi: 10.2166/wh.2024.291.

[56] Q. Xie, J. Tan, H. Li, X. Xu, G. Huang, and Z. Huo, "Assessment of the environmental health and human exposure risk of emerging contaminants in groundwater of a typical agricultural irrigation area in the North China Plain," *J. Hazard. Mater.*, vol. 495, p. 138985, Sep. 2025, doi: 10.1016/J.JHAZMAT.2025.138985.

[57] I. R. S. Salami, N. A. Thufailah, N. Fahimah, and D. Roosmini, "Health risk assessment of physicochemical and heavy metals exposures of the usage of shallow groundwater located at the proximity to Citarum River, Indonesia," *Case Studies in Chemical and Environmental Engineering*, vol. 11, p. 101153, Jun. 2025, doi: 10.1016/J.CSCEE.2025.101153.

[58] T. Zhu, Q. Wu, S. Gao, J. Zeng, K. Linghu, and X. Zhang, "Comparative assessment of health risks and water quality of groundwater in urban and rural Guiyang, Southwest China: Insights from PMF and Monte Carlo Simulation," *Ecotoxicol. Environ. Saf.*, vol. 299, p. 118359, Jul. 2025, doi: 10.1016/J.ECOENV.2025.118359.