FLUORIDE

Quarterly Journal of The International Society for Fluoride Research Inc.

Construction of Artificial Intelligence Driven Public Health Monitoring System for Fluoride: An Exposure Risk Pre-alarm Model Based on Internet Social Behavior Data

Unique digital address (Digital object identifier [DOI] equivalent): https://www.fluorideresearch.online/epub/files/376.pdf

Feiyang SHEN[1], Tao LIU[1*], Ranjit SINGH[2**]

1 Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China

2 Department of Botany, D.S.N. (P.G.), College, Unnao, 209801, (U.P.), India

*** Corresponding author:**
* Dr. Tao Liu, Hangzhou Dianzi University, Hangzhou, Zhejiang, China
Email: lcxn46@163.com

** Dr. Ranjit Singh, Department of Botany, D.S.N. (P.G.), College, Unnao, (U.P.), India
Email: biogreenr@gmail.com

Submitted: 2025 May 21
Accepted: 2025 Aug 11
Published as e376: 2025 Aug 13

ABSTRACT

Purpose: This study focuses on developing an artificial intelligence-driven public health monitoring system for fluoride and constructs a pre-warning risk model based on internet social behavior data.

Methods: Through multi-channel collection of relevant data, text mining technologies and data analysis algorithms were used to analyze the key risk factors associated with fluoride exposure. These factors spanned geographic regions, lifestyle habits, consumption patterns, and other critical variables. Based on this, a predictive early-warning model was developed using an information gain-based feature selection algorithm and a support vector machine (SVM) algorithm. Additionally, corresponding system architecture and functional modules were designed to support the model.

Results: During the experimental validation phase, the data set was divided by cross-validation approach. Model performance assessed based on the accuracy, recall, precision, and the F1-score. The results demonstrate that the model achieved an accuracy of 83%, a recall of 78%, the precision of 80%, and the F1-score of 79%.

Conclusions: The research shows that the model exhibits high reliability and practicability, enabling the effective identification of fluoride exposure risks. It offers innovative and effective technical support for public health monitoring systems.

Keywords: *Fluoride; Public health monitoring; Artificial intelligence; Risk pre-alarm*

INTRODUCTION

Fluoride (F), a chemical substance prevalent in natural environment and various aspects of human life, has significant implications for public health. Adequate F intake effectively prevents dental caries and support bone health. However, excessive consumption may lead to serious health issues, such as dental and skeletal fluorosis.[1] With the acceleration of industrialization and the changes of people's lifestyle, the sources and exposure routes of F are increasingly complex and diverse, which undoubtedly poses a more severe challenge to public health monitoring.[2]

Traditional public health monitoring approaches of F mainly rely on the field sampling and laboratory testing. Such applications are not only expensive, time-consuming and laborious, but also have relatively limited monitoring scope, so it is difficult to achieve real-time and dynamic monitoring and pre-alarm of F exposure risks.[3] With the rapid development of information technology, the social behavior data of the Internet presents an explosive growth trend. These data cover all aspects of people's daily life and contain rich health-related information, which provides a brand-new data source and research approaches for F exposure risk monitoring.[4]

With its advanced capabilities in data mining, analysis, and predictive modeling, AI technology has achieved remarkable results across numerous fields.[5,6] Introducing AI technology in the field of F public health monitoring, and building an pre-alarm model of exposure risk with the help of Internet social behavior data. It is expected to break through the limitations of traditional monitoring approaches and realize accurate identification and pre-alarm of F exposure risk.[7]

Based on this, the purpose of this study is to develop an AI-driven public health monitoring system for F exposure and to conduct an in-depth exploration of risk early-warning model based on internet social behavior data. By conducting a comprehensive and systematic analysis of F exposure risk factors, the AI algorithm model is meticulously designed and optimized to enable efficient monitoring and precise early warning of F exposure risks. Simultaneously, this study will rigorously validate the model to ensure its reliability and practical applicability. The anticipated findings are expected to introduce innovative methodologies and actionable insights for F-related public health monitoring. Ultimately, this research aims to advance public health standards and contribute meaningfully to the field.

MATERIAL AND METHODS

RISK FACTORS OF FLUORIDE EXPOSURE

Accurately identifying relevant risk factors is critical when constructing an early-warning model for F exposure risk. With the help of internet social behavior data, this study assesses the key risk factors through a series of analysis processes.

First, data collection is carried out, and text data and consumption records related to F are obtained from various sources, i.e., social media platforms, health forums, e-commerce site reviews and other channels. The collected data are preprocessed, including removing duplicate information, correcting spelling mistakes, unifying data formats, etc., to improve data quality. After pretreatment, text mining technologies and data analysis algorithm are used to identify risk factors. The study found that factors such as region, living habits and consumption behavior are closely associated to F exposure risk (Table 1). For instance, in certain regions, geological conditions lead to elevated F levels in water, exposing local residents to health risks through drinking water[8]. In terms of living habits, frequent use of toothpaste and drinking specific brands of F drinks may increase F intake[9]. From the perspective of consumption behavior, the high frequency of purchasing some fluorine-containing products reflects the individual's potential F exposure risk.

Table 1 presents different categories of factors and their specific contents, which is helpful to clearly understand the relationship between each factor and F exposure risk, and provide a basis for the subsequent development of exposure risk pre-alarm model.

CONSTRUCTION OF AI DRIVEN EXPOSURE RISK PRE-ALARM MODEL

After identifying the risk factors of F exposure, establishing an accurate and effective early-warning model is essential for public health monitoring. This study designed a comprehensive and efficient exposure risk early-warning model leveraging AI technology. Figure 1 depicts the workflow, encompassing data preprocessing, feature selection, model training, and optimization. The overall model framework adopts hierarchical architecture, which is divided into data input layer, feature processing layer, model operation layer and result output layer. The data input layer is responsible for collecting and integrating the preprocessed internet social behavior data, covering all kinds of risk factor data obtained from the previous analysis. The feature processing layer deeply digs and refines the input data and transforms it into a feature vector that can be directly applied by the model. The model operation layer uses a specific algorithm to analyze and calculate the feature vector to predict the exposure risk of F. The result output layer presents the data results in an intuitive and easy-to-understand form, such as risk grading, which provides a clear basis for public health decision-making.

The purpose of feature selection is to select the responses that play a key role in F exposure risk prediction from different original sources, so as to improve the efficiency and accuracy of the model. In this study, the information gain assesses the importance of each feature to the classification task by measuring its influence on the information entropy of the data set. For a data set D, assuming its category attribute is C, the information gain A of a feature IG(A) is calculated as given equation, as stated-

$$IG(A)=H(C)-\sum_{v\in Values(A)}\frac{|D_v|}{|D|}H(C|A=v) \quad (1)$$

Among them, H(C) is the information entropy of category C, reflecting the uncertainty of category distribution; H(C|A=v) is the conditional entropy of category C, when the value of feature A is v; $|D_v|$ is the size of the data subset, when A is set to |D|, and D is the size of the data set EE. By calculating the information gain of each feature and selecting the feature with higher information gain as the model input. The interference of redundant features are effectively reduced.

Table 1. Statistics factors related to F-exposure risk

Factor category	Specific factor	Description	High-risk group (%)	Low-risk group (%)	Risk coefficient
Region	High-F areas	Groundwater F levels above average due to geology	60	40	1.5
	Areas near industrial pollution	Elevated environmental F from industrial emissions	55	45	1.3
Living habits	Frequency of F toothpaste use	Weekly usage frequency correlates with F exposure	-	-	-
	Daily use	Use F toothpaste at least once daily	70	30	1.8
	3–6 Time/week	Use F toothpaste 3–6 times weekly	45	55	1.1
	<3 Time/week	Use F toothpaste fewer than 3 times weekly	20	80	0.6
Consumption behavior	Number of F product purchases	Monthly count of F-containing consumer goods	-	-	-
	≥5 Items/month	Purchase 5 or more F products monthly	65	35	1.6
	2–4 Items/month	Purchase 2–4 F products monthly	35	65	0.9
	<2 Items/month	Purchase fewer than 2 F products monthly	15	85	0.4
Health habits	Frequency of drinking specific fluoridated beverages	Weekly consumption frequency of specific fluoridated drinks	-	-	-
	≥4 Times/week	Drink specific fluoridated beverages 4 or more times weekly	58	42	1.4
	1–3 Times/week	Drink specific fluoridated beverages 1–3 times weekly	30	70	0.8
	Rarely	Drink specific fluoridated beverages less than once monthly	18	82	0.5

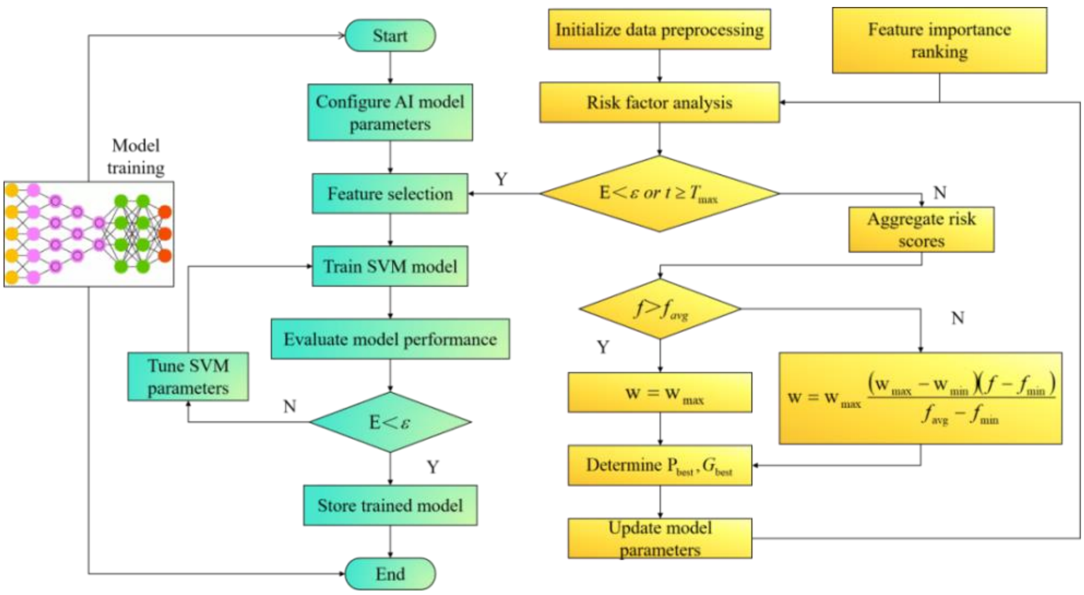

Figure 1. Construction and optimization of F-exposure risk pre-alarm model

In this study, SVM algorithm is selected to build pre-alarm model. SVM is a powerful binary classification model, which can separate different data points by finding the optimal hyperplane. For nonlinear separable data, design the data to high-dimensional space through kernel function, to find the appropriate hyperplane. In the pre-alarm scenario of F exposure risk, it divided into two categories, such as high and low risk. Let the training data set be:

$$\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\} \tag{2}$$

Where x_i is the feature vector determined by the feature selection algorithm, and the corresponding risk categories are as give:

$$y_i\in\{-1,1\}) \tag{3}$$

The goal of SVM is find as a hyperplane:

$$w^Tx+b=0 \tag{4}$$

Next, the interval between two types of data points and hyperplane is maximized. Following optimization problems are solved as:

$$\min_{w,b}\frac{1}{2}||\omega||^2+C\sum_{i=1}^{n}\xi \tag{5}$$

$$\text{s.t.}\ \ y_i(w^Tx_i+b)\geq 1-\xi_i,\ \ \xi_i\geq 0,\ \ i=1,2,\cdots,n \tag{6}$$

Where, C is a penalty parameter, which is used to balance the relationship between interval maximization and classification errors; ξ_i is a slack variable, allowing some data points to have classification errors. By solving the above optimization problems, the optimal ω and b are obtained, as to construct an SVM model that can accurately predict the F exposure risk.

With the cooperation of the above feature selection algorithm and the pre-alarm model algorithm, the AI-driven exposure risk pre-alarm model developed in this study can effectively use the internet social behavior data to accurately predict the F exposure risk and provide core technical support for the F public health monitoring system.

DESIGN OF FLUORIDE PUBLIC HEALTH MONITORING SYSTEM

In order to realize effective monitoring and pre-alarm of F exposure risk, it is very important to design the perfect F public health monitoring system[10]. The system covers two parts, such as system architecture and functional modules.

Fluoride public health monitoring system is designed in a layered architecture, which is mainly divided into data acquisition, data processing, model operation and application display layers. The data acquisition layer is responsible for obtaining F-related data from a variety of data sources, such as the internet social behavior data mentioned above, including social media statements and e-commerce websites records. The data processing layer cleans, transforms and integrates the collected data, removes noise and redundant information, and arranges the data into a format suitable for model processing. The model operation layer uses the AI-driven exposure risk pre-alarm model constructed to analyze and calculate the processed data and get the prediction result of F exposure risk[11]. The application presentation layer presents the model operation results to users in an intuitive and easy-to-understand way, such as generating risk maps and risk grade reports. Each layer cooperates with each other to ensure the efficient operation of the system (Table 2).

TABLE 2. Architecture of the fluoride public health monitoring system

Layer	Function description	Data flow	Technologies involved	Expected processing speed (items/sec)
Data collection layer	Collects F-related data from multiple sources (e.g., online platforms, monitoring stations)	→ Data processing layer	Web crawlers, API calls	100-200
Data processing layer	Cleans, transforms, and integrates data	→ Model operation layer	Data cleaning algorithms, format conversion tools	80-150
Model operation layer	Runs the AI-based risk prediction model	→ Application display layer	AI algorithm framework	50-100
Application display layer	Presents risk results via charts and reports	-	Visualization libraries	-

TABLE 3. Fluoride-related data collection

Data type	Data source	Collection method	Frequency	Purpose
Text data	Social media, health forums	Web crawler	Daily	Analyze public awareness and discussion
Sales data	E-commerce platforms	API access	Twice weekly	Track market circulation of fluoride products
Image data	Product promotional pages	Web crawler	Monthly	Identify product appearance and marketing focus
Monitoring data	Regional water quality stations	Data interface integration	Real-time	Assess environmental F-levels

DATA ACQUISITION MODULE

This module has multi-channel data acquisition function, and can obtain data from the different platforms regularly. According to different data sources, corresponding data collection strategies are adopted, such as collecting text information related to F released by users from social media platforms through web crawlers, and obtaining sales data of fluorine-containing products from e-commerce platforms. The collected data are rich and varied, including text, numbers, images, etc. (Table 3).

RISK ASSESSMENT MODULE

According to the data output from the data processing layer, the module calls the pre-alarm model of the model operation layer to carry out risk assessment. Firstly, feature extraction and matching are carried out on the input data to make it meet the input requirements of the model. Then, the model analyze the data based on the trained parameters, and outputs the evaluation results of F exposure risk, such as the classification of low, medium and high risk.

Pre-alarm release module

When the risk assessment module obtains the risk result, the pre-alarm release module is started immediately. For high-risk situations, timely notify relevant public health departments and people who may be affected by SMS, email, etc. At the same time, pre-alarm information will be released in a prominent position in the system application display interface, providing detailed risk description, impact scope and suggested measures, so that relevant personnel can make timely response decisions. The design of the above system architecture and functional modules, the F public health monitoring system can realize all-round and real-time monitoring and pre-alarm of F exposure risks, and provide strong support for ensuring public health[13].

EXPERIMENTAL VERIFICATION

In order to ensure the reliability and effectiveness of the AI-driven F exposure risk pre-alarm model, it needs to be strictly verified by experiments. In this experiment, the collected data set is divided into training and test set according to the ratio of 7:3 by cross-validation. The training set is used to train the model so that it can learn the characteristics and laws in the data. The test set is used to assess the performance of the model on unknown data. In the training process, the optimal model configuration is found by adjusting model parameters, such as the penalty parameters in SVM algorithm.

Accuracy, Recall, Precision and F1-score are selected as indicators to evaluate the performance of the model. Accuracy measures the proportion of samples correctly predicted by the model to the total number of samples. The recall rate reflects the proportion of the number of positive samples correctly identified by the model to the actual number of positive samples. The accuracy rate represents the proportion of samples predicted as positive examples by the model that are actually positive examples. F1 score is the harmonic average of precision and recall rate, which comprehensively reflects the performance of the model.

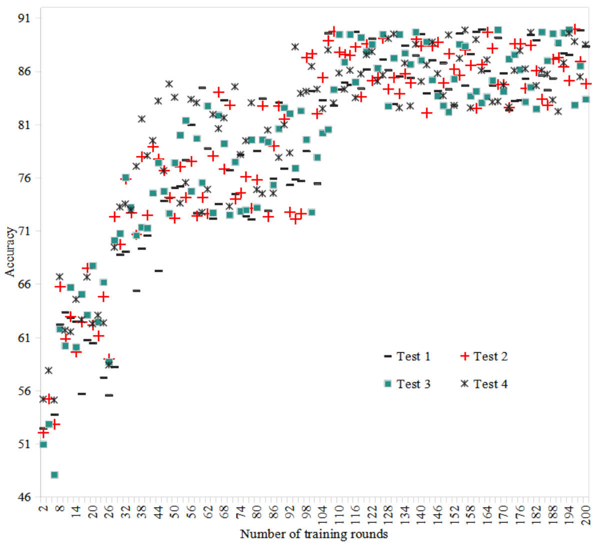

Figure 2. Accuracy vs. number of training epochs

The accuracy of the model on the test set changes with the number of training rounds, as shown in Figure 2. It can be seen that with the increase of training rounds, the accuracy gradually rises and tends to be stable. The final stability is nearly 85%, which shows that the model can correctly predict F exposure risk in most cases.

Figure 3 shows that the changing trend of recall. At the initial stage of training, the recall rate was low, but with the learning of data characteristics by the model, the recall rate gradually increased and finally reached 80%, indicating that the model has a good ability to identify the actual F exposure risk. The precision fluctuated during the training process; it finally stabilized at 82%, which means that most of the samples predicted by the model as high F exposure risk are indeed high-risk samples (Figure 4). Figure 5 shows that the change of F1-score considering the accuracy and recall. The F1-score finally stabilized at nearly 81%, which further proved that the model performed well in overall performance.

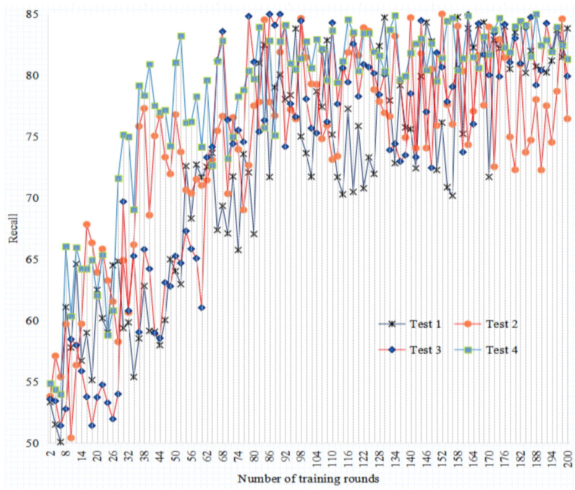

Figure 3. Recall vs. number of training epochs

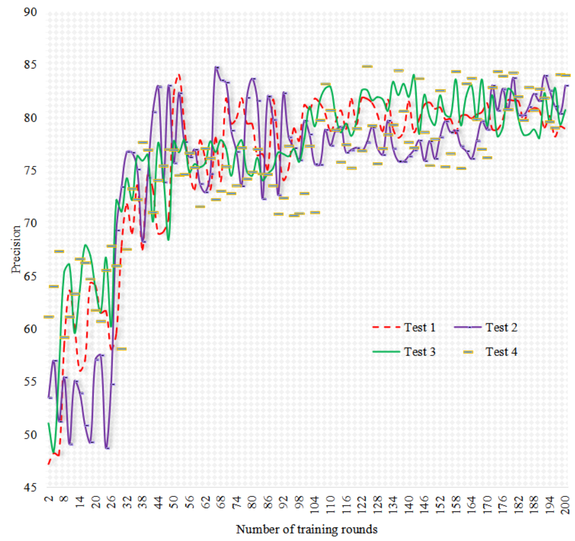

Figure 4. Precision vs. number of training epochs

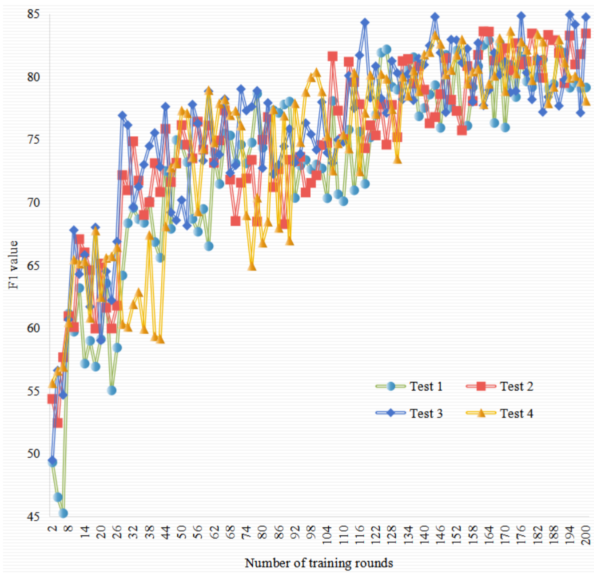

Figure 5. F1-score vs. number of training epochs

In order to compare the performance of each index of the model more intuitively, data summarized in Table 4. From the above experimental results, it can be seen that the pre-alarm model of F exposure risk developed in this study has achieved significant results in accuracy, recall, accuracy and F1-score. It shows that the model can accurately identify F exposure risk, high reliability and practicability, and can provide effective technical support for F public health monitoring[14]. At the same time, through the analysis of the experimental results, it provides a direction for the further optimization of the model. For example, in-view of the large fluctuation of some indicators, the model parameters can be further adjusted or the data processing application can be optimized to improve the stability and generalization ability of the model.

Table 4. Multi-dimensional evaluation of model performance

Metric	Low-risk prediction (%)	High-risk Prediction (%)	Overall (%)	Metric
Accuracy	90	76	83	Accuracy
Recall	85	71	78	Recall
Precision	88	72	80	Precision
F1-Score	86	71	79	F1-Score

CONCLUSION AND RECOMMENDATIONS

The public health monitoring system of F driven by AI and the pre-alarm model of exposure risk based on the social behavior data on the internet were successfully developed. In the risk factor analysis, it is clear that factors such as region, living habits, consumption behavior and health habits are closely associated to F exposure risk. For example, people in specific high-F areas, people around industrial pollution areas, and people who frequently use F toothpaste, F-contaminated water-beverages and buy large quantities of F products have relatively high risk of F exposure.

In the aspect of model construction, with the help of feature selection algorithm and SVM algorithm based on information gain, the prediction performance of the model is effectively improved. The experimental results shows that the accuracy, recall, precision and F1 score of the model reach 83, 78, 80 and 79%, respectively, which shows that the model can accurately identify F exposure risk. From the perspective of system design, the hierarchical monitoring system and each functional module work together to realize the integrated process of data collection, processing, model calculation and result display, which provides an intuitive and effective basis for public health decision-making.

However, there are still some limitations in this study. For example, although the data sources are extensive, there may be incomplete coverage, and more types of data may be considered in the future. At the same time, the stability and generalization ability of the model may need to be further improved in complex and changeable real scenes. Overall, this study offers a contemporary approach to the field of public health monitoring, which holds positive implications for safeguarding public health. Subsequent research can build upon this foundation, with opportunities for further optimization and refinement.

FUNDING

Not applicable

DISCLOSURE OF FINANCIAL AND NON-FINANCIAL RELATIONSHIPS AND ACTIVITIES AND CONFLICTS OF INTEREST

None

REFERENCES

[1] Khattak JA, Farooqi A, Hussain I, Kumar A, Singh CK, Mailloux BJ, et al. Groundwater fluoride across the Punjab plains of Pakistan and India: Distribution and underlying mechanisms. Sci Total Environ 2022;806(1):151353. DOI:10.1016/j.scitotenv.2021.151353

[2] Salunkhe AR, Dudhwadkar S, Raju NP, Tandon S. Public health risk assessment and speciation of air-borne microorganisms in an office building. J Aerosol Sci 2024;179:106362. DOI:10.1016/j.jaerosci.2024.106362

[3] Ahmad M, Jamal A, Tang XW, Al-Sughaiyer MA, Al-Ahmadi HM, Ahmad F. Assessing potable water quality and identifying areas of waterborne diarrheal and fluorosis health risks using spatial interpolation in Peshawar, Pakistan. Water 2020;12(8):2163. DOI:10.3390/w12082163

[4] Hashemi SY, Azari A, Raeesi M, Yaghmaeian K. Application of response surface methodology (RSM) in optimisation of fluoride removal by magnetic chitosan/graphene oxide composite: kinetics and isotherm study. Int J Environ Anal Chem 2023;103:5368-86. DOI:10.1080/03067319.2021.1938021

[5] Garcia ALH, de Souza MR, Picinini J, Soares S, Rohr P, Linden R, et al. Unraveling gene expression and genetic instability in dental fluorosis: investigating the impact of chronic fluoride exposure. Sci Total Environ 2024;906:167393. DOI:10.1016/j.scitotenv.2023.167393.

[6] Jangid NK, Kaliraman S, Singh A, Srivastava A, Srivastava M, Jadoun S, et al. Removal of fluoride ions from water using MgO-based materials with special emphasis on MgO/PPy nanocomposites: A review. J Mol Liquids 2024;399:124473. DOI:10.1016/j.molliq.2024.124473

[7] Khusulio K, Sharma NR, Das IC, Setia RK, Pathak A, Kumar R. Hydrochemical assessment of groundwater with special emphasis on fluoride in parts of Punjab and fluoride prediction using GIS and ML. Environ Earth Sci 2024;83:570. DOI:10.1007/s12665-024-11888-5

[8] Quedi BBB, Ballesteros JrFC, Vilando AC, Lu MC. Recovery of fluoride from wastewater in the form of cryolite granules by fluidized-bed homogeneous crystallization process. J Water Process Engg 2024;66:106063. DOI:10.1016/j.jwpe.2024.106063

[9] Ashrafi SD, Jaafari J, Sattari L, Esmaeilzadeh N, Safari GH. Monitoring and health risk assessment of fluoride in drinking water of East Azerbaijan Province, Iran. Int J Environ Anal Chem 2023;103:1-15. DOI:10.1080/03067319.2020.1849662

[10] Gonzalez Davila O. Dental fluorosis in children from Aguascalientes, Mexico: A persistent public health problem. Water 2021;13(8):1125. DOI:10.3390/w13081125

[11] Medeiros GHA, Soualmia LF, Zanni-Merk C. Towards public health-risk detection and analysis through textual data mining. Procedia Comp Sci 2024;246:3014-23. DOI:10.1016/j.procs.2024.09.370

[12] Paramati SR, Safiullah M, Soytas U. Does geopolitical risk increase carbon emissions and public health risk? Energy Econ 2025;143:108235. DOI:10.1016/j.eneco.2025.108235

[13] Joshi A. Motivating sustainable behaviors by framing biodiversity loss as a public health risk. J Risk Res 2022;25(2):156-175. DOI:10.1080/13669877.2021.1913634

[14] Bixler TS, Collins MR, Mo W. Risk-based public health impact assessment for drinking water contamination emergencies. Sci Total Environ 2024;931:172966. DOI:10.1016/j.scitotenv.2024.172966.