International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN: 2147-6799 www.ijisae.org Original Research Paper

A Comparative Analysis of Machine Learning Approaches for Predicting Dental Fluorosis Risk from Groundwater Fluoride Contamination

Gaurav Saxena[1*], Dr. Priyank Singhal[2]

Submitted: 16/09/2024 Revised: 20/10/2024 Accepted: 30/11/2024

Abstract
Dental fluorosis is a significant public health concern caused by excessive intake of groundwater fluoride contamination, particularly in regions with high natural fluoride deposits. This study systematically explores various Machine Learning (ML) approaches used in predicting dental fluorosis risk, focusing on key techniques such as Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Extreme Gradient Boosting (XGBoost). Following a comprehensive literature search from 2019 to 2024 using databases like Scopus, IEEE, Web of Science, and Springer, this analysis adheres to PRISMA guidelines to ensure methodological rigor. Studies demonstrated promising results, with advanced models like SVM and MLP achieving high prediction accuracy rates of up to 97.4%. The key observation of this study is that these advanced models outperform traditional methods in predicting fluoride contamination risks. However, challenges remain, particularly in developing region-specific models, improving data collection, and addressing computational inefficiencies. This study provides a detailed comparative analysis of these techniques and identifies research gaps, including the need for scalable models and improved real-time prediction for effective public health interventions.

Keywords: *Dental Fluorosis, Groundwater Fluoride Contamination, Machine Learning, Predictive Models, Multilayer Perceptron*

1. Introduction

The element fluorine is never found in nature in its elemental form since it is the most electronegative chemical element [1-3]. As a chemical aggregate of fluorides, it makes up around 0.06 to 0.09% of the Earth's crust and ranks seventeenth among the elements that make up the crust [4]. The most abundant materials are hornblende, fluorapatite, mica, and cryolite. The high levels of fluoride in natural waters are caused by mica minerals, volcanic rocks, and thermal causes [5,6]. The typical concentration of fluoride in surface waters is below 1 mg/L. Such concentrations can reach 20–53 mg/L in hot spring waters or deep underground water sources that have come into touch with fluoride-rich rocks [7-9]. Fluoride is present in the human body, in rocks and soil, in the air and water, and in plants and animals. While teeth and bone growth can benefit from low fluoride intake, fluorosis, bone and teeth loss, and other health issues (e.g., negative effects on children's intellectual development) can result from long-term excessive fluoride absorption [10–13].

[1]*Research Scholar, Department of Computer Science, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India, kumar1.saxena@gmail.com*
[2]*Associate professor, Department of Computer Science, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India, drpriyanksinghal@gmail.com*

There are several potential sources of high fluoride levels in humans, including companies that generate minerals, coal combustion, volcanic ash, brick tea (a type of compressed tea), and water that is polluted with fluoride [14–17]. Although it is most prevalent in areas with limited access to water, fluorosis is quickly becoming an environmental toxicological concern worldwide.

Most of the fluorine that enters the human body comes from fluoride that is naturally present in drinking water. Individuals residing in regions with fluoride concentrations higher than the daily optimum dose are at increased risk of developing endemic fluorosis, a serious public health concern. Food and water are the most common entry points for fluoride to enter in human body [18-21]. Studies show that the majority of people's daily fluoride consumption comes from drinkable water, which is a significant source of soluble fluoride. Ion chromatography, spectrophotometry, an ion-selective electrode, and inductively coupled plasma are some of the chemical analytical tools used to determine the amount of fluoride in water [22–26].

Groundwater fluoride contamination poses significant dental health risks, particularly dental fluorosis, which occurs when excessive fluoride is ingested during the development of teeth [27-30]. Fluoride levels exceeding 1.5 mg/L, commonly found in regions with high natural fluoride deposits, lead to dental fluorosis, manifesting as discoloration and damage to tooth enamel. In mild cases, it appears as white streaks, but in severe cases results in brown, pitted teeth, compromising both aesthetics and enamel strength [31–33]. Children are most vulnerable, especially in fluoride-affected areas where untreated groundwater is a primary drinking source. While fluoride at optimal levels (0.7–1.2 mg/L) prevents dental caries by strengthening enamel, excessive amounts cause long-term damage. Regions with water-stressed conditions, such as parts of Africa and Asia, are particularly impacted, as they often lack resources to mitigate fluoride levels [34-37]. Figure 1 shows different types of fluorosis.

Figure 1: Different types of fluorosis [38]

Dental fluorosis remains a significant public health concern, particularly in regions with high natural fluoride deposits, where excessive fluoride intake can lead to substantial dental damage. This issue underscores the critical need for effective water management and continuous monitoring of fluoride levels to prevent overexposure.

Determining the amount of dental fluorosis requires a thorough clinical examination of the entire oral cavity, with a particular focus on the morphology of the teeth. Appearance is significant in diagnosis. It can vary from minute, barely noticeable small white spots or flecks to much more serious discoloration, pitting, and enamel loss [39-41]. Severity is categorized on a scale from 0, or no fluorosis, to 9, which reflects severe fluorosis using the Thylstrup-Fejerskov index [42]. This index takes into account the size, color, and position of damaged areas on the teeth. A dentist can use some dental imaging techniques, for example, X-rays or radiographs, along with a direct visual inspection to diagnose how severe this condition is or to detect other potential issues in dental health [43].

Dental fluorosis is primarily morphological, but it is painless. However, dental damage can be rehabilitated through restorative procedures for better function and aesthetics in more severe cases. Minor therapies applied in the mild moderate to severe cases of dental fluorosis might be the use of treatment techniques such as teeth whitening or microabrasion [44]. In contrast to microabrasion, which includes the removal of a very thin layer of enamel to improve the look of the teeth, tooth whitening uses bleaching solutions to lighten the afflicted teeth [45,46]. Restoration of both form and function can be necessary in more extensive cases, necessitating restorative dental procedures such as dental bonding, veneers, or crowns. It can be required to extract the tooth and replace it with an implant if the enamel deterioration is significant [47-49].

Treating dental fluorosis rectifies the harm incurred but does not resolve the fundamental issue of excessive fluoride consumption from environmental sources. Consequently, alongside treatment, it is imperative to avert additional excessive fluoride consumption to diminish the prevalence of fluorosis [50-52].

1.1 Machine Learning (ML) for Dental Fluorosis Risk Prediction

ML techniques have also seen extensive application in the field of medicine in the last several years [53–55]. The previous methods are precise and quantitative, but they are often time-consuming, difficult, and costly. To take advantage of ML's cost-effectiveness, chemical measurements or analysis are required. These chemical analytical approaches, however, are becoming less common as a result of advancements in ML [56].

Consequently, researchers, technical workers, and medical personnel must be able to detect fluoride quickly and affordably [57–59]. For applications such as speech recognition, autonomous vehicles, and complex data analysis, ML teaches computers to emulate human cognitive functions [60,61]. Biological nervous systems, such as the brain, inspire Artificial Neural Networks (ANNs), a paradigm for information processing [62]. ANNs represent an information processing framework derived from biological nervous systems, such as the human brain. ANNs are employed in the problem-solving processes analogous to those of the human brain [63]. Support Vector Machine (SVM) is a classification technique grounded in statistical learning theory, relying on two fundamental concepts. It constructs a hyperplane that optimally distinguishes the two classes and maximizes the separation between them [64]. The Naïve Bayes classifier is a method that calculates the likelihood of a new observation fitting into a specified category, utilizing a probability model based on Bayes' theorem [65].

Figure 2 illustrates a general ML model process for dental risk prediction, utilizing patient data from an Electronic Dental Records (EDR) database. The model draws on various categories of patient information, including demographics, medical history, periodontal charting, and dental history. After an initial collection phase, patients with incomplete or missing critical data are excluded, refining the pool of eligible cases. The remaining dataset is divided into training and testing sets, with a focus on numerous unique features. The model is trained using an advanced algorithm, XGBoost, and its performance is assessed through metrics like the Receiver Operating Characteristic (ROC) curve, aiming to accurately predict dental risk outcomes [66].

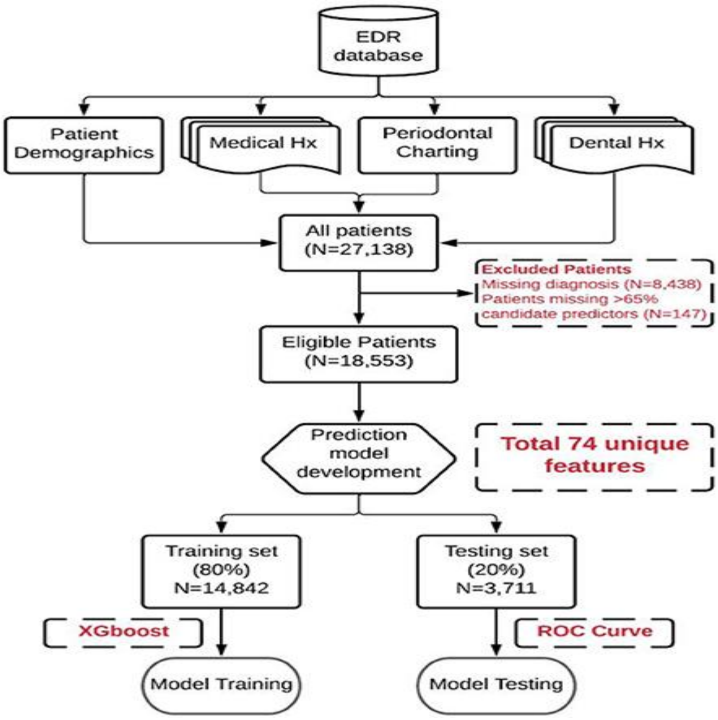

Figure 2: General framework for dental risk prediction using ML [66].

Over the last years, ML has become one of the most promising tools for predictive analytics that promises to open new possibilities in the prediction of the risk for dental diseases, including those that carry an ecological dimension, such as the contamination of groundwater [67]. Dental fluorosis caused by fluoridation during the period of development of teeth is a health problem in all parts of the world, particularly in regions where the content of fluoride in the groundwater is significant [68]. This is a game changer to use ML techniques in the field because by accessing huge databases, ML models can predict more efficiently at a faster speed and on an enormous scale [69]. It is important to consider this, especially for dealing with environmental health issues where one needs to predict the disease burden caused by toxic substances present in the environment. In populations where groundwater is exposed to increased fluoride contamination, ML systems can successfully target geographic regions at high risk of fluorosis and estimate prevalence levels, enabling effective public health campaigns [70,71].

The objective of this study is to review and evaluate various ML techniques used to predict dental fluorosis risks associated with groundwater fluoride contamination. It aims to assess the predictive accuracy, scalability, and utility of different ML algorithms in identifying high-risk populations and regions. Additionally, the paper explores challenges related to data collection and processing, especially in resource-limited settings, and examines how predictive models can contribute to public health interventions and policies focused on reducing fluoride exposure and preventing dental fluorosis.

The contribution of this study lies in identifying gaps in current research, such as the lack of region-specific predictive models, challenges in data collection, and limitations in existing model performance. The review outlines avenues of future research work including more advanced and refined ML models concerning the non-linear relationship between fluoride exposure and health outcomes, better techniques for collecting data, and real-time monitoring data for sharp predictions. These make the applications of public health, policy-making, and preventive strategies more feasible for areas affected by fluoride.

2. Literature review

In this section, various studies and research focused on Predicting Dental Fluorosis Risk from Groundwater Fluoride Contamination using ML are discussed.

Mohammed et al., (2024) [72] analyzed groundwater contamination through supervised and unsupervised machine algorithms, namely, Multilayer Perceptron Artificial Neural Networks (MLP-ANN) and Principal Component Analysis (PCA), to identify sources of fluoride problem as well as the non-carcinogenic health risk due to nitrate and fluoride contaminant exposure. All the produced models were tested for performance concerning their Coefficients of Determination (R^2) as well as the Root Mean Square Error (RMSE). The model exhibited results to be excellent by ensuring an R^2 higher than 97% along with a maximum RMSE of less than 4% within both training and validation sessions. Similarly, **Mallik et al., (2023) [73]** estimated the performance of hybrid Monte Carlo ML (MC-ML) models through the hazard quotients used for health hazards. Physical parameters, along with nitrates, were determined through the analysis of 32 groundwater samples. The hybrid Monte Carlo-Deep Neural Network (MC-DNN) model was considered to perform better in comparison with other models in both test and train phases from the output of the prediction model based on error and performance metrics. Before this, **Wongkhuenkaew et al., (2023) [74]** proposed an automated approach for image segmentation and classification of dental fluorosis images. While classifying the features, researchers considered the fuzzy k-nearest neighbor method, and for optimizing the cluster number cuckoo search algorithm was accounted for. Finally, the criterion was developed based on ratios of opaque-to-brown pixels for the four-class classification of fluorosis. Experimental results carried out on 128 blind test images reported average pixel accuracy in all four classes of fluorosis as 92.24%. Apart from that, **Nafouanti et al., (2023) [75]** proposed a new estimation method of fluoride in the groundwater: Hybrid Random Forest Linear Model (HRFLM). The capability of multiple models in predicting subsurface water fluoride contamination was estimated by taking 202 samples of groundwater. Based on the outcome, the hybrid HRFLM attained an accuracy rate of 95%. In addition to that, the Area Under the Curve (AUC) value of the HRFLM was observed to be 0.98. However, **Sarkar et al., (2023) [76]** examined the relationship between fluoride and the tectonic framework using a multi-model strategy. Based on fluoride data taken all around India, researchers have used three models based on ML to forecast groundwater fluoride levels. With a 93% accuracy rate, the random forest model was the best. Along with "depth to the water table," tectonics were determined to be a significant factor.

Cao et al., (2022) [77] employed a technique called Siamese Network-Based Transfer Learning (SNTL) to determine the chances of dangerous compounds being present in groundwater above a certain level using sparse and unequal data. To remove the detrimental impacts of class-imbalanced data on prediction model performance, SNTL drastically cut down the amount of training data that was needed. Compared to other benchmark models, SNTL models offered better sensitivity and specificity (approximately 80%). Moreover, **Ling et al., (2022) [78]** evaluated groundwater samples in Pakistan at a spatial resolution of 250 m using ML algorithms to determine the probability that fluoride concentrations would exceed 1.5 mg/L. Factors such as weather, soil, lithology, terrain, and vegetation cover were found to be strong indicators of elevated fluoride levels in groundwater. With an AUC of 0.92 on non-modeling test data, a random forest model was found to exhibit excellent performance. Groundwater in the Thar Desert has been reported to have the greatest quantities of fluoride.

Similarly, **Pang et al., (2021) [79]** developed a novel model that incorporated both genetic and environmental variables into teenage dental risk prediction using an ML algorithm. Cohort 1 (the training cohort) yielded an AUC of 0.78 for the researchers' random forest-based caries risk prediction model. Using cohort 2, researchers confirmed that this caries risk prediction model could discriminate and calibrate. With an AUC of 0.73 in the testing group, the caries risk prediction model demonstrated strong discriminatory power. Additionally, **Podgorski et al., (2018) [80]** modeled 12,600 groundwater fluoride concentrations across India using a random forest ML technique, with the use of spatially continuous predictor factors mainly about geology, climate, and soil parameters. Fluoride surpassing 1.5 mg/L at 1 km resolution has been accurately predicted globally using this method. Predictions estimate the overall number of individuals in danger of fluorosis from groundwater fluoride at about 120 million, or about 9% of the population. Lastly, **Sajedi et al., (2018) [81]** formulated an innovative approach for assessing the risk of nitrate pollution in groundwater by combining chemical and statistical analyses for an arid environment. Three ML models were employed to assess the chances of groundwater pollution occurrence. The models were validated using AUC, with values exceeding 80 percent. The results demonstrated that the accuracy of the three models varied between 0.81 and 0.87; therefore, all models were considered suitable for the ensemble modeling procedure.

ML has revolutionized environmental risk prediction models, particularly for fluoride contamination-related dental health risks such as dental fluorosis. Traditional detection methods are often labor-intensive and costly, but ML models offer more efficient, scalable, and cost-effective alternatives. By analyzing large datasets, algorithms like random forests, neural networks, and decision trees can predict fluoride contamination in groundwater and assess health risks with high accuracy. These models integrate complex environmental data, enabling the identification of high-risk populations and informing public health interventions. Additionally, ML-driven models reduce the reliance on expensive chemical analyses, making fluoride monitoring more accessible, especially in resource-limited regions, while improving the speed and precision of risk assessments.

3. Review Methodology

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement is utilized as a set of criteria for screening and refining research. A detailed literature assessment on this phenomenon was conducted using the SCOPUS database, which includes records published from 2019 to 2024. This review includes all records categorized as articles, journals, and publications from the Scopus database. Table 1 shows the keywords that are used in the Scopus database to find the papers related to the topic:

Table 1: Searching Keywords

Source: Authors own elaboration

Databases	Keyword used
Scopus	TITLE-ABS-KEY ("Machine Learning" AND "Dental Fluorosis" OR "Groundwater" AND "Risk Prediction" OR "Environmental Health" OR "Predictive Analytics" OR "Public Health" OR "Water Quality" OR "Fluorosis Prediction Models")

In literature analysis, only records categorized as articles are subjected to further analysis and assessment, as they are considered for evaluation purposes. The below-presented table 2 exclusively examined records that met the specified inclusion and exclusion criteria:

Table 2: The criteria for determining what is Included and Excluded

Source: Authors own elaboration

Criterion	Inclusion	Exclusion
Keywords	Records conferring the relationship between machine learning (ML) approaches for predicting dental fluorosis risk from groundwater fluoride contamination.	Records excluded in which variables have no relation.
Type of Literature	Journals, Review Articles	Book, book series, book chapter.
Language	English	Other than English
Timeframe	Concerning 2019-2024	<2019
Category	Open Access	Paid Access

3.1 Prisma Model

3.2 The PRISMA statement assists the reviews "preferred reporting items for systematic reviews and meta-analyses". PRISMA demonstrates the review's caliber and enables readers to grasp its advantages and disadvantages as well as replicate review strategies.

Figure 3 depicts the comprehensive PRISMA flow diagram outlining the process of study identification, screening, eligibility assessment, and inclusion. The process begins with the retrieval of 17,400 studies from previous reviews, alongside new records sourced from databases such as IEEE, Web of Science, Scopus, ResearchGate, ScienceDirect, and Springer, collectively yielding 5,432 records. Following this, an automated deduplication process resulted in the exclusion of 16,968 records deemed ineligible. The remaining 432 records were subjected to an initial screening phase, during which 16 records were excluded based on the inclusion timeframe of 2019-2024. Subsequently, 416 reports were identified for full-text retrieval, of which 401 were further evaluated for eligibility against predefined criteria. At this stage, 11 reports were excluded due to their publication status, while 1 report was excluded due to language incompatibility. Ultimately, 30 studies met all inclusion criteria and were incorporated into the systematic review, ensuring the relevance and rigour of the evidence base for the research question. Figure 4 depicts the number of selected papers from the top 6 research areas. Environmental Science takes the lead with 291 documents, which presents it as an important branch in ecological and sustainability studies. In any general development concerning economics and industry, engineering makes 64 documents outlining its involvement in such fields as infrastructure, technology and industrial development. Moreover, 63 documents from the Social Sciences stream recommend the analysis of people, their society and culture, as well as the environment.

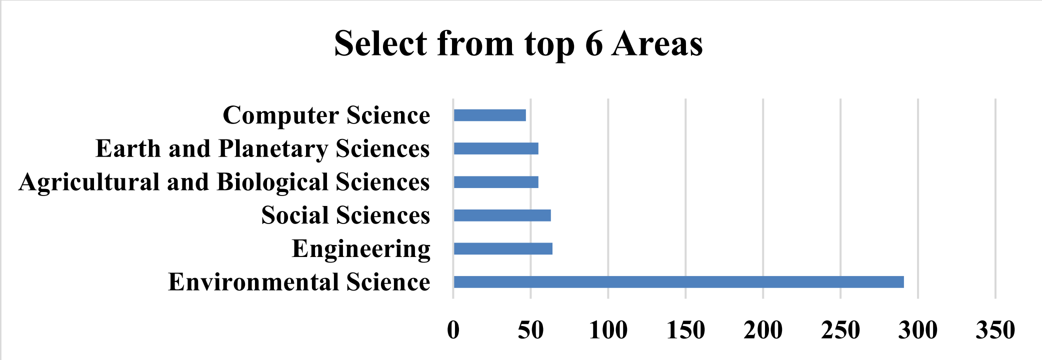

Figure 3: Selected papers from the top 6 Areas.
Source: Authors own elaboration

Agricultural and Biological Sciences and Earth and Planetary Sciences contain 55 each that represent science practice and geological study respectively. There is technology, data processing and computational research in Computer Science with 47 documents.

Figure 5 illustrates the distribution of various keywords relevant to research in the field of water resource management using advanced techniques. "Machine Learning" leads with 320 documents, reflecting its significant role in predictive modeling, optimization, and data analysis across various applications. "Water Quality" follows closely with 295 documents, emphasizing the focus on assessing and maintaining water standards through advanced methodologies. "Ground Water" is featured in 137 documents, highlighting its importance in the study of underground water reserves and their sustainable management. Lastly, "Prediction" appears in 102 documents, showcasing its relevance in forecasting trends and outcomes in water resource studies. This data underscores the significant intersection of ML and water quality research, with a particular focus on groundwater studies and predictive capabilities.

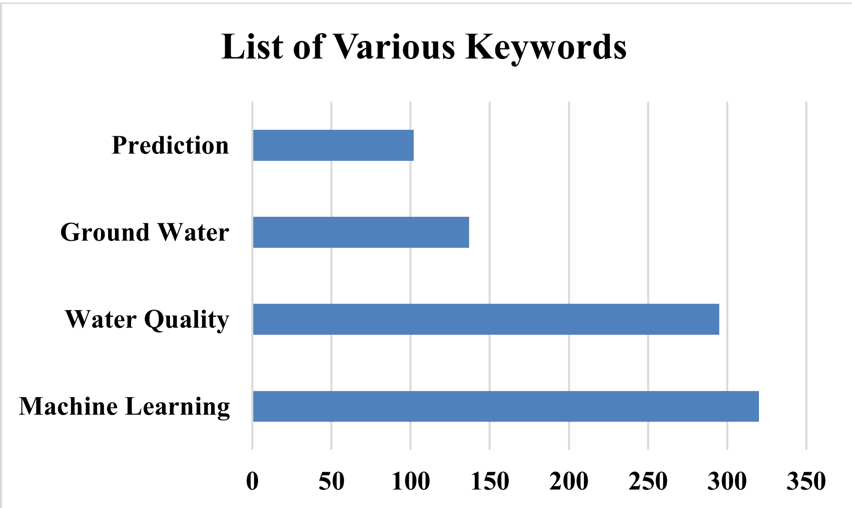

Figure 4: List of various Keywords
Source: Authors own elaboration

Figure 6 offers valuable information regarding the years in which the papers were published. For this systematic literature review, the search terms were selected specifically to explore risk prediction models for dental fluorosis from groundwater fluoride contamination using ML and deep learning techniques.

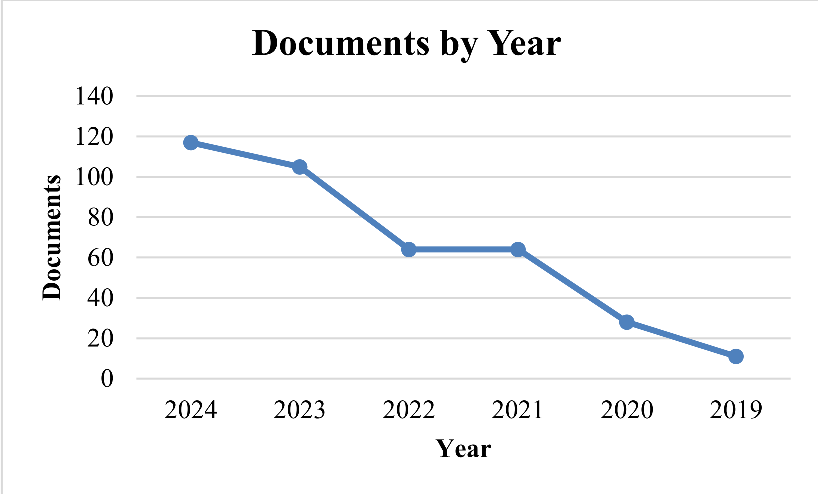

Figure 5: Trend in published papers.
Source: Authors own elaboration

3.3 Research questions

The following are the research questions (Q1 - Q3) derived from the objectives of this SLR:

Q1: What machine learning techniques have been used to predict dental fluorosis risk from groundwater fluoride contamination?

Q2: What are the main environmental and chemical factors considered in the predictive models for dental fluorosis?

Q3: How can machine learning models help improve public health efforts and prevent dental fluorosis risk?

4. Machine learning techniques to predict dental fluorosis risk

ML techniques have shown significant potential in predicting dental fluorosis risk due to their ability to analyze complex datasets. **Ogwo et al., (2024) [82]** applied various ML models in the prediction of dental risk in young adults, including Generalized Boosting Machines (GBM), Least Absolute Shrinkage and Selection Operator (LASSO) regression, XGBOOST, and Negative Binomial (NegGLM). In the evaluation of the prediction analysis, it was found that LASSO was the best model, with a minimum RMSE of 0.70 and a maximum R^2 of 0.44. Accuracy, precision, recall, and AUC for classification results were 83.7%, 85.9%, 93.1%, and 68.2%, respectively. Similarly, **Bomfim et al., (2024) [83]** utilized an XGboost, logistic regression, and decision tree for the prevalence of untreated dental caries for teenagers. Eight major predictor factors were cross-validated against the epidemiological base to classify the patients affected by untreated dental care. The analysis was undertaken using the R software tool, and XGboost carried out better performance than a decision tree algorithm at a level of 84 % AUC. Various ML algorithms are used by **Çiftçi et al., (2024) [84]** to predict factors determining adult oral health. 25% of the dataset is used for testing, and the rest 75% was applied to train the model. The F1-score achieved was 96%, accuracy was 95.8%, with recall and precision rate being 96%. The output was also concluded that the highest success rate belongs to the Multilayer Perceptron model. In addition, **Gad et al., (2024) [85]** evaluated the quality of the groundwater through various indexing methods like Drinking Water Quality Index (DWQI). Through the simulation model using a Recurrent Neural Network (RNN), the indices regarding health and water quality, both for training and testing, were well predicted. With an R^2 value of 0.96 accompanied by an RMSE value of 2.73, the model proved to be a phenomenal fit for the training set. Alongside this, **Khusulio et al., (2024) [86]** applied complex techniques, comprising Geographic Information Systems (GIS), along with the ML algorithms to predict the fluoride contamination and compute the Water Quality Index (WQI). The model constructed using the Random Forest (RF) technique yielded a high predictive capability concerning fluoride contamination with AUCs of 81%, 82%, and 94% in the districts of Muktsar, Bathinda, and Moga, respectively.

Moreover, **Gad et al., (2024) [87]** explored the potential of using artificial intelligence in predicting dental health based on variables such as health behaviors and orthodontic treatment. The Exploratory Data Analysis (EDA) methods were employed first, followed by the Artificial Intelligence (AI) algorithms. The cross-validation method was used to test the accuracy and generalization ability of the prediction models. In both online datasets, the study was able to achieve nearly 90% accuracy of predictions. Additionally, **Mohseni et al., (2024) [88]** calculated the Weighted Arithmetic-Water Quality Index (WA-WQI) to assess groundwater quality. Four ML models were created to forecast the WQI in Ujjain, Madhya Pradesh, India. All ML model findings were validated using the AUC curve for optimal model selection and achieved the highest value of 0.9048. Likewise, **Venkatesh et al., (2024) [89]** developed a Water Quality Index (WQI) using mathematical and ML analysis. Results from ML experiments showed that a model combining Ridge Regression (RR) and Random Forest (RF) performed better than other ML models, with a validation accuracy of 99.9 percent.

Similarly, **Ibrahim et al., (2023) [90]** assessed the quality of the Ground Water (GW) for agricultural land in Egypt's Western Desert, specifically in El Kharga Oasis. In order to forecast eight different Irrigation Water Quality Indices (IWQIs), two ML models were created. According to the results, the majority of the GW samples were rated as excellent (100%), good (57.85%), or excellent to good (65.71%). Further,

Paneerselvam et al., (2023) [91] designed the Agriculture Suitability Index (ASI) model and combined it with the Radial Basic Function (RBF) model to determine the most important parameter in the chemical equilibrium of groundwater for agricultural use. The groundwater's nitrate concentration degraded its overall quality, and the results demonstrated that more than 85 percent of sample sites were acceptable for human consumption. According to the RBF model, the R^2 values for summer were 0.84, and for winter, they were 0.85. Furthermore, **Hussein et al., (2023) [92]** evaluated the water's potability and irrigation potential using the WQI. In order to choose the most effective classifier from among the available ML methods, the classification learners were applied to both normalized and raw data sets. According to the findings, SVM outperformed all other classifiers for both raw and normalized data. With raw data, the anticipated accuracy level was 90.8%, while with normalized data, it was 89.2%. Along with this, **Sadegh et al., (2022) [93]** explored the possible risk factors for dental caries and lifestyle choices in children younger than five. Using two evaluation approaches (K-fold and Leave-One-Out) and 10 ML modeling techniques, a highly accurate classification model was built to forecast the likelihood of caries. With a 97.4 percent success rate, MLP and RF were the most accurate of the ML algorithms put into practice. Compared to Extreme Gradient Boosting, SVM utilizing Radial Basis Function (RBF) Kernel achieved a higher accuracy of 97.4%. Additionally, **Cao et al., (2022) [94]** employed the artificial neural network to forecast the contamination of geogenic groundwater with fluoride in China. The results indicated that the model's accuracy and AUC in the test dataset were 80.5% and 0.86%, respectively. The most effective predictors were climatic variables.

Awais et al., (2021) [95] developed a new approach to evaluate the potential for nitrate pollution of groundwater. In order to determine the possibility of groundwater pollution, three ML models were employed. According to the findings, the models' accuracy ranged from 0.82 to 0.87. Moreover, **Tran et al., (2021) [96]** analyzed the impact of several parameters on groundwater salinity prediction using ML techniques. The results demonstrated that the Cat Boost Regression (CBR) model outperformed the others on both the training and testing datasets, with $R^2 = 0.999$ and RMSE = 29.90 and $R^2 = 0.84$ and RMSE = 205.96, respectively. Likewise, **Aysegül et al., (2021) [97]** investigated a more economical method that made use of ML technology. In addition, the Simulated Annealing (SA) search method was used to choose the input parameters. With correlation coefficient values of 0.731, the subgroup chosen by the SA algorithm was highly correlated. Experiments showed that by reducing feature size significantly, learning models' prediction performance was enhanced using carefully chosen feature subsets. Alongside this, **Leong et al., (2021) [98]** estimated WQI via the use of a Least Square-Support Vector Machine (LS-SVM). The model's output was further examined using the R^2. Results showed that the polynomial kernel function, with an R^2 of 0.8796, was the most effective kernel function for the support vector machine model. However, **Khosravi et al., (2020) [99]** evaluated the utility of new data mining techniques for predicting fluoride concentrations in groundwater, including lazy Learners Instance-Based K-Nearest Neighbors (IBKNN). Two portions of the complete dataset were extracted, one for model training (the calibration) and another for model evaluation (verification). The IBK model was found to be more effective than competing models in predicting the presence of fluoride contamination. Lastly, **Chen et al., (2019) [100]** proposed a novel hybrid integration approach for assessing groundwater potential that combined the Fisher's Linear Discriminant Analysis (FLDA) with Rotation Forest (RFLDA) and Bagging (BFLDA) ensembles. To assess the components' predictive ability, researchers utilized the classifier attribute evaluation approach on the FLDA model. The BFLDA model outperformed the RFLDA and FLDA models in terms of prediction competence, with a value of 95%.

Table 3 presents a summary of various ML and hybrid techniques applied by different authors for risk prediction of Dental Fluorosis.

Table 3: Performance analysis of various ML and hybrid technologies used for risk prediction of Dental Fluorosis

Author	**Technique Used**	**Accuracy**	**Dataset Used**	**Research Gaps**
Ogwo et al. (2024) [82]	LASSO	83.7%	Iowa Fluoride dataset	The lack of predictive models for dental caries in young adults using diverse, longitudinal datasets limits generalizability across demographics.
Bomfim et al. (2024) [83]	XGBoost,	84%,	Epidemiological survey dataset	Models lack generalizability across diverse populations, limiting broader applicability.
Çiftçi et al. (2024) [84]	Multilayer Perceptron	95.8%	Oral dataset	Lack of focus on long-term follow-up and clinical applicability
Gad et al. (2024) [85]	RNN	~90%	Primary dataset	Insufficient use of integrated ML and GIS models for predicting groundwater quality risks across diverse geochemical regions.
Khusulio et al. (2024) [86]	RF	94%	-	Lack of comprehensive predictive modeling combining GIS and ML for fluoride contamination risk assessment.

Gad et al. (2024) [87]	EDA	~90%	Primary dataset	Lack of a dedicated ML model specifically for predicting dental fluorosis risk based on fluoride exposure patterns.
Mohseni et al. (2024) [88]	XG-Boost	90.48	-	Limited exploration of model-specific performance, interpretability, and adaptability in predicting dental health issues.
Venkatesh et al. (2024) [89]	RF	96.9%	Primary dataset	Lack of comprehensive exploration into the long-term predictive accuracy and interpretability of ML models.
Ibrahim et al. (2023) [90]	SVM	67.85%	Primary dataset	Inadequate focus on long-term impact assessment of groundwater contamination on agricultural practices.
Paneerselvam et al. (2023) [91]	RBF	84%	Agriculture suitability dataset	Limited seasonal focus, lacking extended time frame analysis
Hussein et al. (2023) [92]	SVM	90.8%	Primary dataset	Limited exploration of ML model optimization for water quality index prediction, especially in varying geographic regions
Sadegh et al. (2022) [93]	SVM with RBF,	97.4%,	Primary dataset	Limited focus on regional dietary patterns and their effect on caries
Cao et al. (2022) [94]	ANN	80.5%	-	Limited high-resolution models combining socio-environmental factors for precise fluoride risk assessment across China
Awais et al. (2021) [95]	DRASTIC	87%	Primary dataset	Specific to nitrate, lacking cross-contaminant comparison
Tran et al. (2021) [96]	CBR	84%	Groundwater salinity dataset	Limited evaluation of ML models specifically for groundwater salinity prediction in complex, multi-layer coastal aquifer systems
Aysegül et al. (2021) [97]	SA	73.1%	Primary dataset	Limited research exists on cost-effective, rapid prediction of groundwater fluoride levels using optimized ML models in geogenic fluoride-prone areas.
Leong et al. (2021) [98]	LS-SVM	87.96%	-	Need for more accurate and standardized ML models to predict the WQI efficiently across diverse environmental conditions.
Khosravi et al. (2020) [99]	IBK	-	Primary dataset	Specific to fluoride, the lack of a broader contaminant focus
Chen et al. (2019) [100]	FLDA with RFLDA, BFLDA	95%	Groundwater potential dataset	Insufficient investigation of hybrid machine learning models for mapping groundwater potential.

Figure 7 illustrates the accuracy of different ML techniques. SVM with RBF Kernel achieves the highest accuracy at 97.40%, followed by Multilayer Perceptron at 95.80%. XGBoost and LASSO show competitive performance with 84% and 83.70%, respectively. This comparison highlights that SVM with RBF Kernel and Multilayer Perceptron are the most effective models.

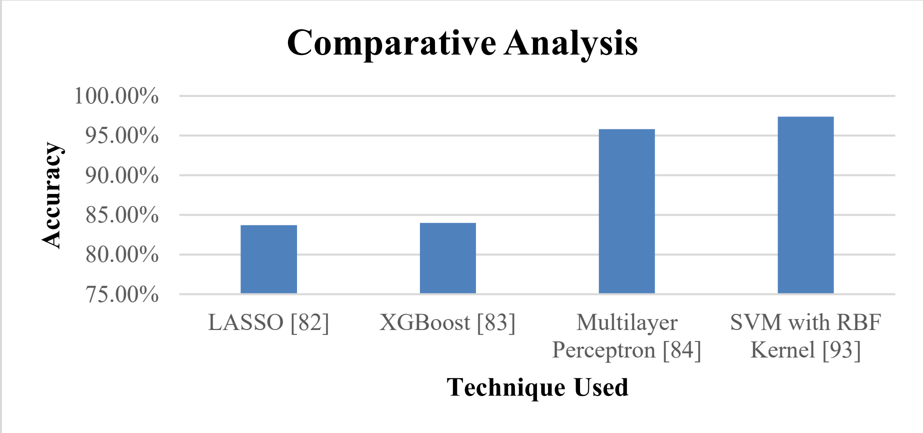

Figure 6: Comparative Analysis of various machine learning techniques

5. Discussion and Comparison

This section performs a statistical analysis of the ML methods discussed in previous sections. It compares techniques and accuracies of models for predicting dental fluorosis risk, highlighting methods like SVM, RF, MLP, XGBoost, and hybrid models. Furthermore, the proposed AQs proposed in Section 3 are called to present some analytical reports as follows:

Q1: What machine learning techniques have been used to predict dental fluorosis risk from groundwater fluoride contamination?

The machine learning techniques used to predict dental fluorosis risk from groundwater fluoride contamination include LASSO, GBM, NegGLM, XGBoost, Decision Tree, and Logistic Regression. Advanced models like MLP, Random Forest, and SVM with RBF Kernel were also applied, along with Simulated Annealing for feature selection. Among these, MLP and SVM with RBF Kernel achieved the highest accuracy at 97.4%, while XGBoost and LASSO showed competitive performance with accuracies of 94.9% and 83.7%, respectively. These approaches highlight diverse and effective methods for dental fluorosis prediction.

Q2: What are the main environmental and chemical factors considered in the predictive models for dental fluorosis?

The main environmental and chemical factors considered in predictive models for dental fluorosis include fluoride concentration in groundwater, the presence of fluoride-rich minerals such as mica and volcanic rocks, and the characteristics of water sources, such as whether they come from deep wells or hot springs. Another factor is the availability of clean water and rural/urban settings in socioeconomic and demographic characteristics. Dietary factors like fluoride ingestion through food and beverages and climate-related factors, especially in arid areas where the dependence on groundwater is higher, are also important. These are considered in ML models to better predict the risk associated with dental fluorosis.

Q3: How can machine learning models help improve public health efforts and prevent dental fluorosis risk?

Machine learning models improve public health efforts significantly and aid risk prevention of dental fluorosis by allowing more accurate and early detection of populations at risk. Analysis of big datasets like fluoride concentration of groundwater, demographic information, and environmental factors for the identification of existing patterns that would not be exhibited by the traditional method can be done. Advanced techniques like XGBoost, LASSO, and SVM can predict risk factors of dental fluorosis more accurately, so targeted interventions could be done in those regions where fluoride is over the safe limits. In addition, feature selection techniques such as Simulated Annealing optimize the input variables by taking only the relevant ones. This leads to an efficient use of resources for the most vulnerable populations while guiding public health policies towards preventive measures such as the improvement in water management and monitoring of fluoride. In this regard, ML enhances the decision-making process so that timely and effective action to mitigate the risk of dental fluorosis can be performed.

6. Conclusion and Future Scope

This study presents a comparative analysis of machine learning approaches applied to predict the risk of dental fluorosis caused by fluoride contamination in groundwater, based on 30 rigorously selected papers published between 2019 and 2024, with strict adherence to PRISMA guidelines. The study used to screen the paper includes the type of document, language, and publication year that would eventually lead to the retention of only studies that have been instrumental in explaining ML's role in the risk prediction of dental fluorosis. In the traditional models, XGBoost, and Decision Tree have been used to draw a solution for Q1 as well as the respective accuracy of each. The advanced model applied in the previous sections is SVM with RBF Kernel and Multilayer Perceptron, from which it has been evident that 97.4% prediction accuracy was obtained. Q2 focuses on incorporating some factors related to environmental and chemical parameters, such as fluoride concentration in the groundwater characteristics, along with a few socioeconomic-related variables, to enhance predictions. From Q3, the report highlighted gaps in the research work: missing in the study region-specific predictive models, limited coverage of socio-economic influences on the outcome, and problems in data collection and processing procedures in resource-poor settings. Future research should address these shortcomings by developing regionally adapted models, improving collection methods of data, and real-time prediction models towards the enhancement of public health interventions. Models of ML at present have shown huge advancement in the task of dental fluorosis prediction but efforts should be towards such refinement of models that are effective for real-world scalable efficiency and diversity in practical scenarios of public health.

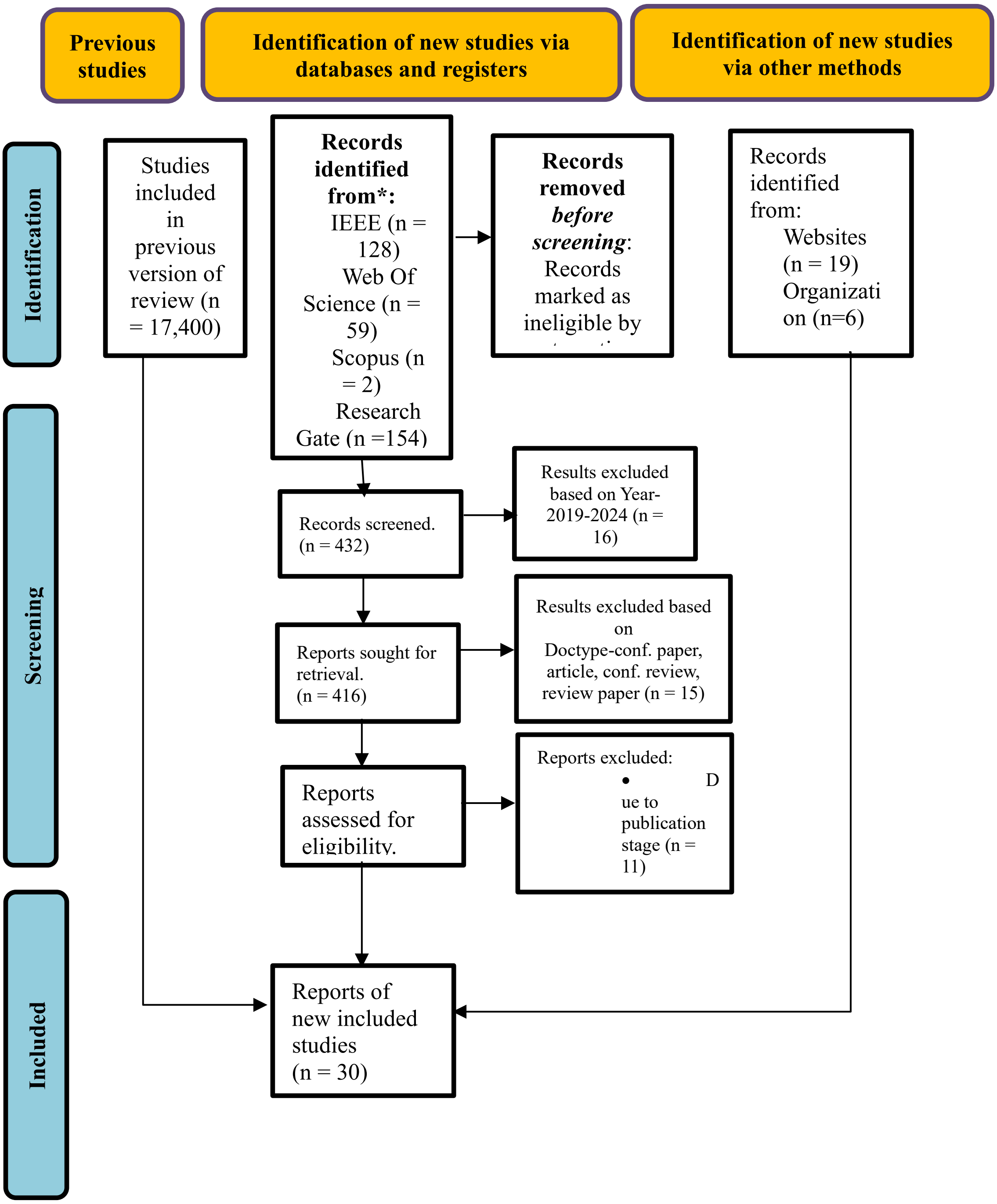

Figure 7: Prisma Model

References

[1] Deshmukh, A. N., P. M. Wadaskar, and D. B. Malpe. "Fluorine in the environment: A review." *Gondwana Geol. Mag* 9 (1995): 1-20.

[2] Dehnen, Stefanie, Laurel L. Schafer, Thomas Lectka, and Antonio Togni. "Fluorine: a very special element and its very special impacts on chemistry." *Inorganic Chemistry* 60, no. 23 (2021): 17419-17425.

[3] World Health Organization. *Fluorine and fluorides*. World Health Organization, 1984.

[4] World Health Organization. "Fluorides and oral health: report of a WHO Expert Committee on Oral Health Status and Fluoride Use [meeting held in Geneva from 22 to 28 November 1993]." (1994).

[5] Brindha, K., and L. Elango. "Fluoride in groundwater: causes, implications and mitigation measures." *Fluoride properties, applications and environmental management* 1 (2011): 111-136.

[6] Yeşilnacar, Mehmet İrfan, Ayşegül Demir Yetiş, Çoruh Türksel Dülgergil, Mustafa Kumral, Ayşe Dilek Atasoy, Tuba Rastgeldi Doğan, Sadık İsmail Tekiner, Ibrahim Bayhan, and Murat Aydoğdu. "Geomedical assessment of an area having high-fluoride groundwater in southeastern Turkey." *Environmental Earth Sciences* 75 (2016): 1-14.

[7] Alloway, Brian, José A. Centeno, Robert B. Finkelman, Ron Fuge, Ulf Lindh, and Pauline Smedley. *Essentials of medical geology*. Edited by Olle Selinus. Academic Press, 2005.

[8] Msonda, K. W. M., W. R. L. Masamba, and E. Fabiano. "A study of fluoride groundwater occurrence in Nathenje, Lilongwe, Malawi." *Physics and Chemistry of the Earth, Parts A/B/C* 32, no. 15-18 (2007): 1178-1184.

[9] Yetiş, Recep, Ayşe Dilek Atasoy, Ayşegül Demir Yetiş, and Mehmet İrfan Yeşilnacar. "Hydrogeochemical characteristics and quality assessment of groundwater in Balikligol Basin, Sanliurfa, Turkey." *Environmental Earth Sciences* 78 (2019): 1-17.

[10] Featherstone, John DB. "Prevention and reversal of dental caries: role of low-level fluoride." *Community Dentistry and Oral Epidemiology* 27, no. 1 (1999): 31-40.

[11] Singh, Gayatri, Babita Kumari, Geetgovind Sinam, Navin Kumar, and Shekhar Mallick. "Fluoride distribution and contamination in the water, soil and plants continuum and its remedial technologies, an Indian perspective–a review." *Environmental Pollution* 239 (2018): 95-108.

[12] O Mullane, D. M., R. J. Baez, S. Jones, M. A. Lennon, P. E. Petersen, A. J. Rugg-Gunn, H. Whelton, and Gary M. Whitford. "Fluoride and oral health." *Community Dental Health* 33, no. 2 (2016): 69-99.

[13] Petersen, Poul Erik, and Hiroshi Ogawa. "Prevention of dental caries through the use of fluoride–the WHO approach." *Community Dent Health* 33, no. 2 (2016): 66-8.

[14] Yetis, Aysegul Demir, Mehmet Irfan Yesilnacar, and Musa Atas. "A machine learning approach to dental fluorosis classification." *Arabian Journal of Geosciences* 14, no. 2 (2021): 95.

[15] Kisku, Ganesh Chandra, and Pokhraj Sahu. "Fluoride contamination and health effects: an Indian scenario." *Environmental Concerns and Sustainable Development: Volume 1: Air, Water and Energy Resources* (2020): 213-233.

[16] Bryson, Christopher. *The fluoride deception*. Seven Stories Press, 2004.

[17] Ananian, Arbi, Benjamin H. Solomowitz, and Ingrid A. Dowrich. "Fluoride: a controversy revisited." *New York State Dental Journal* 72, no. 3 (2006): 14.

[18] Chen, Huafang, Mei Yan, Xuefei Yang, Zhong Chen, Guangan Wang, Dietrich Schmidt-Vogt, Yunfen Xu, and Jianchu Xu. "Spatial distribution and temporal variation of high fluoride contents in groundwater and prevalence of fluorosis in humans in Yuanmou County, Southwest China." *Journal of Hazardous Materials* 235 (2012): 201-209.

[19] Gupta, Sonam, Nivedita Mishra, Ankit Kumar, and Akhilesh Kumar Yadav. "Fluoride Pollution in Subsurface Water: Challenges and Opportunities." *Advanced Treatment Technologies for Fluoride Removal in Water: Water Purification* (2024): 19-39.

[20] Islam, Md Shajedul, and M. G. Mostafa. "Meta-analysis and risk assessment of fluoride contamination in groundwater." *Water Environment Research* 93, no. 8 (2021): 1194-1216.

[21] Mukherjee, Indrani, and Umesh Kumar Singh. "Groundwater fluoride contamination, probable release, and containment mechanisms: a review on the Indian context." *Environmental geochemistry and health* 40, no. 6 (2018): 2259-2301.

[22] Ghosh, Aniruddha, Kakali Mukherjee, Sumanta K. Ghosh, and Bidyut Saha. "Sources and toxicity of fluoride in the environment." *Research on Chemical Intermediates* 39 (2013): 2881-2915.

[23] Fawell, John Kirtley. *Fluoride in drinking water*. World Health Organization, 2006.

[24] Srivastava, Sakshi, and S. J. S. Flora. "Fluoride in drinking water and skeletal fluorosis: a review of the global impact." *Current Environmental Health Reports* 7 (2020): 140-146.

[25] Kabir, Humayun, Ashok Kumar Gupta, and Subhasish Tripathy. "Fluoride and human health: Systematic appraisal of sources, exposures, metabolism, and toxicity." *Critical Reviews in Environmental Science and Technology* 50, no. 11 (2020): 1116-1193.

[26] Jha, Sunil Kumar, Vinay Kumar Mishra, Dinesh Kumar Sharma, and Thukkaram Damodaran. "Fluoride in the environment and

its metabolism in humans." *Reviews of Environmental Contamination and Toxicology Volume 211* (2011): 121-142.

[27] Viswanathan, Gopalan, A. Jaswanth, S. Gopalakrishnan, and G. Aditya. "Determining the optimal fluoride concentration in drinking water for fluoride endemic regions in South India." *Science of the Total Environment* 407, no. 20 (2009): 5298-5307.

[28] Gevera, Patrick, Hassina Mouri, and Godfrey Maronga. "Occurrence of fluorosis in a population living in a high-fluoride groundwater area: Nakuru area in the Central Kenyan Rift Valley." *Environmental geochemistry and health* 41 (2019): 829-840.

[29] Senthilkumar, Mohanavelu, Biswarup Mohapatra, Devadasan Gnanasundar, and Sourabh Gupta. "Identifying fluoride endemic areas and exposure pathways for assessment of non-carcinogenic human health risk associated with groundwater fluoride for Gujarat state, India." *Environmental Science and Pollution Research* 28, no. 36 (2021): 50188-50203.

[30] Guan, Zhizhong, Lihua Wang, and Dianjun Sun. "Endemic fluorosis." *Endemic Disease in China* (2019): 61-96.

[31] Peckham, Stephen, and Niyi Awofeso. "Water fluoridation: a critical review of the physiological effects of ingested fluoride as a public health intervention." *The Scientific World Journal* 2014, no. 1 (2014): 293019.

[32] Dhar, Vineet, and Maheep Bhatnagar. "Physiology and toxicity of fluoride." *Indian Journal of Dental Research* 20, no. 3 (2009): 350-355.

[33] Buzalaf, Marília Afonso Rabelo, and Gary Milton Whitford. "Fluoride metabolism." *Fluoride and the oral environment* 22 (2011): 20-36.

[34] Mendes, Ana Luiza G., Mariele S. Nascimento, Rochele S. Picoloto, Erico MM Flores, and Paola A. Mello. "A sample preparation method for fluoride detection by potentiometry with ion-selective electrode in medicinal plants." *Journal of Fluorine Chemistry* 231 (2020): 109459.

[35] Dhillon, Ankita, Manjula Nair, and Dinesh Kumar. "Analytical methods for determination and sensing of fluoride in biotic and abiotic sources: a review." *Analytical Methods* 8, no. 27 (2016): 5338-5352.

[36] Tolkou, Athanasia K., Natalia Manousi, George A. Zachariadis, Ioannis A. Katsoyiannis, and Eleni A. Deliyanni. "Recently developed adsorbing materials for fluoride removal from water and fluoride analytical determination techniques: A review." *Sustainability* 13, no. 13 (2021): 7061.

[37] Michalski, Rajmund. "Ion chromatography as a reference method for determination of inorganic ions in water and wastewater." *Critical Reviews in Analytical Chemistry* 36, no. 2 (2006): 107-127.

[38] Srivastava, Rangoli, Pradeep Tangade, and Surbhi Priyadarshi. "Managing dental fluorosis: A guide to improving the appearance of the teeth." *Archives of Dental Research* 13, no. 1 (2023): 25-29.

[39] Kerketta, Anjali, Harmanpreet Singh Kapoor, and Prafulla Kumar Sahoo. "Groundwater fluoride prediction modeling using physicochemical parameters in Punjab, India: a machine-learning approach." *Frontiers in Soil Science* 4 (2024): 1407502.

[40] Majanga, Vincent, and Serestina Viriri. "A survey of dental caries segmentation and detection techniques." *The Scientific World Journal* 2022, no. 1 (2022): 8415705.

[41] Al-Namankany, Abeer. "Influence of artificial intelligence-driven diagnostic tools on treatment decision-making in early childhood caries: a systematic review of accuracy and clinical outcomes." *Dentistry Journal* 11, no. 9 (2023): 214.

[42] Chirani, R. Arbab, and H. Foray. "Fluorose dentaire: diagnostic étiologique." *Archives de pédiatrie* 12, no. 3 (2005): 284-287.

[43] Revelo-Mejía, Inés A., Arturo Hardisson, Carmen Rubio, Ángel J. Gutiérrez, and Soraya Paz. "Dental fluorosis: the risk of misdiagnosis—a Review." *Biological Trace Element Research* 199 (2021): 1762-1770.

[44] Kashirtsev, Filipp, John Tressel, and Daniel Fried. "Dehydration imaging of dental fluorosis at 1950 nm." In *Lasers in Dentistry XXVIII*, vol. 11942, pp. 47-52. SPIE, 2022.

[45] Greenwall, Linda. "Tooth Whitening, the Microabrasion Technique, and White Spot Eradication." In *Tooth Whitening Techniques*, pp. 173-186. CRC Press, 2017.

[46] Majeed, Abdul, Imran Farooq, Sias R. Grobler, and R. J. Rossouw. "Tooth-bleaching: a review of the efficacy and adverse effects of various tooth whitening products." *J Coll Physicians Surg Pak* 25, no. 12 (2015): 891-6.

[47] Shahroom, Nor Syakirah Binti, Geo Mani, and Mahesh Ramakrishnan. "Interventions in management of dental fluorosis, an endemic disease: A systematic review." *Journal of Family Medicine and Primary Care* 8, no. 10 (2019): 3108-3113.

[48] Zotti, Francesca, Luca Albertini, Nicolò Tomizioli, Giorgia Capocasale, and Massimo Albanese. "Resin infiltration in dental fluorosis treatment—1-year follow-up." *Medicina* 57, no. 1 (2020): 22.

[49] Rathi, B. Senthil, P. Senthil Kumar, Gayathri Rangasamy, Michael Badawi, and Tejraj M. Aminabhavi. "Membrane-based removal of fluoride from groundwater." *Chemical Engineering Journal* (2024): 150880.

[50] Greenwall-Cohen, J., L. Greenwall, V. Haywood, and K. Harley. "Tooth whitening for the under-18-year-old patient." *British Dental Journal* 225, no. 1 (2018): 19-26.

[51] Pandey, Pallavi, Afroz Alam Ansari, Preeti Moda, and Madhulika Yadav. "Enamel

microabrasion for aesthetic management of dental fluorosis." *Case Reports* 2013 (2013): bcr2013010517.

[52] Tajuddin, Nazrizawati Ahmad, Ezza Faridha Binti Sokeri, Nur Alyaa Kamal, and Mustapha Dib. "Fluoride removal in drinking water using layered double hydroxide materials: preparation, characterization and the current perspective on IR4. 0 technologies." *Journal of Environmental Chemical Engineering* 11, no. 3 (2023): 110305.

[53] Groznik, Vida, Matej Guid, Aleksander Sadikov, Martin Možina, Dejan Georgiev, Veronika Kragelj, Samo Ribarič, Zvezdan Pirtošek, and Ivan Bratko. "Elicitation of neurological knowledge with argument-based machine learning." *Artificial intelligence in medicine* 57, no. 2 (2013): 133-144.

[54] Nakano Y, Takeshita T, Kamio N, Shiota S, Shibata SN, Yoneda M, Hirofuji T, Yoshihisa Y (2014) Supervised machine learning based classification of oral malodor based on the microbiota in saliva samples. Artif Intell Med 60:97–101

[55] Pérez, Noel Pérez, Miguel A. Guevara López, Augusto Silva, and Isabel Ramos. "Improving the Mann–Whitney statistical test for feature selection: An approach in breast cancer diagnosis on mammography." *Artificial intelligence in medicine* 63, no. 1 (2015): 19-31.

[56] Lin, Jia-yu, Hong-tao Liu, and Jun Zhang. "Recent advances in the application of machine learning methods to improve identification of the microplastics in the environment." *Chemosphere* 307 (2022): 136092.

[57] Eyke, Natalie S., Brent A. Koscher, and Klavs F. Jensen. "Toward machine learning-enhanced high-throughput experimentation." *Trends in Chemistry* 3, no. 2 (2021): 120-132.

[58] Aoun, Antoine, Farah Darwiche, Sibelle Al Hayek, and Jacqueline Doumit. "The fluoride debate: the pros and cons of fluoridation." *Preventive Nutrition and Food Science* 23, no. 3 (2018): 171.

[59] Whelton, H. P., A. J. Spencer, L. G. Do, and A. J. Rugg-Gunn. "Fluoride revolution and dental caries: evolution of policies for global use." *Journal of Dental Research* 98, no. 8 (2019): 837-846.

[60] Sanuade, Oluseun A., Amjed M. Hassan, Adesoji O. Akanji, Abayomi A. Olaojo, Michael A. Oladunjoye, and Abdulazeez Abdulraheem. "New empirical equation to estimate the soil moisture content based on thermal properties using machine learning techniques." *Arabian journal of geosciences* 13 (2020): 1-14.

[61] Srinivas, T., G. Aditya Sai, and R. Mahalaxmi. "A comprehensive survey of techniques, applications, and challenges in deep learning: A revolution in machine learning." *International Journal of Mechanical Engineering* 7, no. 5 (2022): 286-296.

[62] Yang, Q., Z. Hou, Y. Wang, Y. Zhao, and J. Delgado. "A comparative study of shallow groundwater level simulation with WA–ANN and ITS model in a coastal island of south China." *Arabian Journal of Geosciences* 8 (2015): 6583-6593.

[63] Khashei-Siuki A, Sarbazi M (2015) Evaluation of ANFIS, ANN, and geostatistical models to the spatial distribution of groundwater quality (case study: Mashhad plain in Iran). Arab J Geosci 8:903–912

[64] Gasmi, Anis, Cécile Gomez, Hédi Zouari, Antoine Masse, and Danielle Ducrot. "PCA and SVM as geo-computational methods for geological mapping in the southern of Tunisia, using ASTER remote sensing data set." *Arabian Journal of Geosciences* 9 (2016): 1-12.

[65] Tsangaratos P, Ilia I (2016) Comparison of logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: the influence of models complexity and training dataset size. Catena 145:164–179

[66] Patel, Jay S., Chang Su, Marisol Tellez, Jasim M. Albandar, Rishi Rao, Vishnu Iyer, Evan Shi, and Huanmei Wu. "Developing and testing a prediction model for periodontal disease using machine learning and big electronic dental record data." *Frontiers in Artificial Intelligence* 5 (2022): 979525.

[67] Zhong, Shifa, Kai Zhang, Majid Bagheri, Joel G. Burken, April Gu, Baikun Li, Xingmao Ma et al. "Machine learning: new ideas and tools in environmental science and engineering." *Environmental science & technology* 55, no. 19 (2021): 12741-12754.

[68] Vithanage, Meththika, and Prosun Bhattacharya. "Fluoride in drinking water: health effects and remediation." *CO2 sequestration, biofuels and depollution* (2015): 105-151.

[69] Zhou, Lina, Shimei Pan, Jianwu Wang, and Athanasios V. Vasilakos. "Machine learning on big data: Opportunities and challenges." *Neurocomputing* 237 (2017): 350-361.

[70] Gansky, Stuart A., and Sarah Shafik. "At the crossroads of oral health inequities and precision public health." *Journal of Public Health Dentistry* 80 (2020): S14-S22.

[71] Arudra, Annepu, Dankan Gowda, Parashuram Shankar Vadar, Mahadeo Ramchandra Jadhav, and Mandeep Kaur. "Predictive Modeling of Dental Health Outcomes Based on Fluoride Concentrations using AI." In *2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, pp. 1-7. IEEE, 2023.

[72] Mohammed, M. A. A., A. Mohamed, N. P. Szabó, and P. Szűcs. "Development of machine-learning-based models for identifying the sources of nitrate and fluoride in

groundwater and predicting their human health risks." *International Journal of Energy and Water Resources* 8, no. 2 (2024): 161-180.

[73] Mallik, Santanu, Saikat Das, Abhigyan Chakraborty, Umesh Mishra, Swapan Talukdar, Somnath Bera, and G. V. Ramana. "Prediction of non-carcinogenic health risk using Hybrid Monte Carlo-machine learning approach." *Human and Ecological Risk Assessment: An International Journal* 29, no. 3-4 (2023): 777-800.

[74] Wongkhuenkaew, Ritipong, Sansanee Auephanwiriyakul, Nipon Theera-Umpon, Kasemsit Teeyapan, and Uklid Yeesarapat. "Fuzzy K-nearest neighbor based dental fluorosis classification using multi-prototype unsupervised possibilistic fuzzy clustering via cuckoo search algorithm." *International Journal of Environmental Research and Public Health* 20, no. 4 (2023): 3394.

[75] Nafouanti, Mouigni Baraka, Junxia Li, Edwin E. Nyakilla, Grant Charles Mwakipunda, and Alvin Mulashani. "A novel hybrid random forest linear model approach for forecasting groundwater fluoride contamination." *Environmental Science and Pollution Research* 30, no. 17 (2023): 50661-50674.

[76] Sarkar, Soumyajit, Abhijit Mukherjee, Madhumita Chakraborty, Md Tahseen Quamar, Srimanti Duttagupta, and Animesh Bhattacharya. "Prediction of elevated groundwater fluoride across India using multi-model approach: insights on the influence of geologic and environmental factors." *Environmental Science and Pollution Research* 30, no. 11 (2023): 31998-32013.

[77] Cao, Hailong, Xianjun Xie, Jianbo Shi, Guibin Jiang, and Yanxin Wang. "Siamese network-based transfer learning model to predict geogenic contaminated groundwaters." *Environmental Science & Technology* 56, no. 15 (2022): 11071-11079.

[78] Ling, Yuya, Joel Podgorski, Muhammad Sadiq, Hifza Rasheed, Syed Ali Musstjab Akber Shah Eqani, and Michael Berg. "Monitoring and prediction of high fluoride concentrations in groundwater in Pakistan." *Science of the Total Environment* 839 (2022): 156058.

[79] Pang, Liangyue, Ketian Wang, Ye Tao, Qinghui Zhi, Jianming Zhang, and Huancai Lin. "A new model for caries risk prediction in teenagers using a machine learning algorithm based on environmental and genetic factors." *Frontiers in Genetics* 12 (2021): 636867.

[80] Podgorski, Joel E., Pawan Labhasetwar, Dipankar Saha, and Michael Berg. "Prediction modeling and mapping of groundwater fluoride contamination throughout India." *Environmental Science & Technology* 52, no. 17 (2018): 9889-9898.

[81] Sajedi-Hosseini, Farzaneh, Arash Malekian, Bahram Choubin, Omid Rahmati, Sabrina Cipullo, Frederic Coulon, and Biswajeet Pradhan. "A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination." *Science of the total environment* 644 (2018): 954-962.

[82] Ogwo, Chukwuebuka, Grant Brown, John Warren, Daniel Caplan, and Steven Levy. "Predicting dental caries outcomes in young adults using machine learning approach." *BMC Oral Health* 24, no. 1 (2024): 529.

[83] Bomfim, Rafael Aiello. "Machine learning to predict untreated dental caries in adolescents." *BMC Oral Health* 24, no. 1 (2024): 316.

[84] Çiftçi, Burak Tunahan, and Firdevs Aşantoğrol. "Utilization of machine learning models in predicting caries risk groups and oral health-related risk factors in adults." *BMC Oral Health* 24, no. 1 (2024): 430.

[85] Gad, Mohamed, Aissam Gaagai, Asmaa A. Agrama, Walaa FM El-Fiqy, Mohamed Hamdy Eid, Péter Szűcs, Salah Elsayed et al. "Comprehensive evaluation and prediction of groundwater quality and risk indices using quantitative approaches, multivariate analysis, and machine learning models: An exploratory study." *Heliyon* 10, no. 17 (2024).

[86] Khusulio, K., Neeta Raj Sharma, Iswar Chandra Das, R. K. Setia, Akhilesh Pathak, and Rohan Kumar. "Hydrochemical assessment of groundwater with special emphasis on fluoride in parts of Punjab and fluoride prediction using GIS and ML." *Environmental Earth Sciences* 83, no. 19 (2024): 570.

[87] Gad, Asser Mohamed. "A Comprehensive Hybrid Approach for Predicting Oral Health Status Based on Different Factors Through New Artificial Intelligence Techniques." *Egyptian Dental Journal* 70, no. 4 (2024): 3063-3073.

[88] Mohseni, Usman, Chaitanya B. Pande, Subodh Chandra Pal, and Fahad Alshehri. "Prediction of weighted arithmetic water quality index for urban water quality using ensemble machine learning model." *Chemosphere* 352 (2024): 141393.

[89] Venkatesh, Arumugasamy Thangapandian, Sujatha Rajkumar, and Uma Shankar Masilamani. "Assessment of groundwater quality using water quality index, multivariate statistical analysis and machine learning techniques in the vicinity of an open dumping yard." *Environment, Development and Sustainability* (2024): 1-26.

[90] Ibrahim, Hekmat, Zaher Mundher Yaseen, Miklas Scholz, Mumtaz Ali, Mohamed Gad, Salah Elsayed, Mosaad Khadr et al. "Evaluation and prediction of groundwater quality for irrigation using an integrated water quality indices, machine learning models and

GIS approaches A representative case study." *Water* 15, no. 4 (2023): 694.
[91] Paneerselvam, Balamurugan, Nagavinothini Ravichandran, Peiyue Li, Maciej Thomas, Warit Charoenlerkthawin, and Butsawan Bidorn. "Machine learning approach to evaluate the groundwater quality and human health risk for sustainable drinking and irrigation purposes in South India." *Chemosphere* 336 (2023): 139228.
[92] Hussein, Enas E., Muhammad Yousuf Jat Baloch, Anam Nigar, Hussain F. Abualkhair, Faisal Khaled Aldawood, and Elsayed Tageldin. "Machine learning algorithms for predicting the water quality index." *Water* 15, no. 20 (2023): 3540.
[93] Sadegh-Zadeh, Seyed-Ali, Ali Rahmani Qeranqayeh, Elhadj Benkhalifa, David Dyke, Lynda Taylor, and Mahshid Bagheri. "Dental caries risk assessment in children 5 years old and under via machine learning." *Dentistry Journal* 10, no. 9 (2022): 164.
[94] Cao, Hailong, Xianjun Xie, Yanxin Wang, and Hongxing Liu. "Predicting geogenic groundwater fluoride contamination throughout China." *Journal of Environmental Sciences* 115 (2022): 140-148.
[95] Awais, Muhammad, Bilal Aslam, Ahsen Maqsoom, Umer Khalil, Fahim Ullah, Sheheryar Azam, and Muhammad Imran. "Assessing nitrate contamination risks in groundwater: a machine learning approach." *Applied Sciences* 11, no. 21 (2021): 10034.
[96] Tran, Dang An, Maki Tsujimura, Nam Thang Ha, Doan Van Binh, Thanh Duc Dang, Quang-Van Doan, Dieu Tien Bui, Trieu Anh Ngoc, Pham Thi Bich Thuc, and Tien Dat Pham. "Evaluating the predictive power of different machine learning algorithms for groundwater salinity prediction of multi-layer coastal aquifers in the Mekong Delta, Vietnam." *Ecological Indicators* 127 (2021): 107790.
[97] Aysegül, Musa Atas Mehmet Irfan Yesilnacar, and Demir Yetis. "Novel machine learning techniques-based hybrid models (LR-KNN-ANN and SVM) in prediction of dental fluorosis in groundwater." (2021).
[98] Leong, Wei Cong, Alireza Bahadori, Jie Zhang, and Zainal Ahmad. "Prediction of water quality index (WQI) using support vector machine (SVM) and least square-support vector machine (LS-SVM)." *International Journal of River Basin Management* 19, no. 2 (2021): 149-156.
[99] Khosravi, Khabat, Rahim Barzegar, Shaghayegh Miraki, Jan Adamowski, Prasad Daggupati, Mohammad Reza Alizadeh, Binh Thai Pham, and Mohammad Taghi Alami. "Stochastic modeling of groundwater fluoride contamination: Introducing lazy learners." *Groundwater* 58, no. 5 (2020): 723-734.
[100] Chen, Wei, Biswajeet Pradhan, Shaojun Li, Himan Shahabi, Hossein Mojaddadi Rizeei, Enke Hou, and Shengquan Wang. "Novel hybrid integration approach of bagging-based fisher’s linear discriminant function for groundwater potential analysis." *Natural Resources Research* 28 (2019): 1239-1258.