



# OPEN Machine learning models for water safety enhancement

Fatemeh Ranjbar<sup>1</sup>, Hossein Sadeghi<sup>1✉</sup>, Reza Pourimani<sup>1</sup> & Soraya Khanmohammadi<sup>2</sup>

Humans encounter both natural and artificial radiation sources, including cosmic rays, primordial radionuclides, and radiation generated by human activities. These radionuclides can infiltrate the human body through various pathways, potentially leading to cancer and genetic mutations. A study was conducted using random sampling to assess the concentrations of radioactive isotopes and heavy metals in mineral water from Iran, consumable at Arak City. Notably, specific radiation levels of Ra-226 were not detected, whereas the concentrations of Th-232, K-40, and Cs-137 were found to be below the thresholds established by the World Health Organization (WHO). The annual effective doses derived from the consumption of bottled water were significantly lower than the limits set by the United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR), thereby reducing the risk of cancer. Furthermore, heavy metals such as lead and chromium were not present in the samples, thereby contributing to the overall safety of the water. The Machine Learning (ML) models employed in this study provided accurate predictions, ensuring reliability across various demographic groups and reinforcing the robustness of the findings. Overall, the results suggest that consumable mineral water consumption poses minimal health risks.

**Keywords** Mineral water, Water safety, Machine learning, Health risks, Radioactive isotopes, Potential cancer risk

Water is crucial to human nutrition, with approximately 80% of waterborne diseases attributed to contamination<sup>1,2</sup>. To mitigate disease risks and avoid high monitoring costs due to pollutants from population growth, industrial expansion, and waste disposal, it is essential to prioritize the accessibility and affordability of clean drinking water<sup>3</sup>.

The human body is exposed to both natural and artificial radiation. Natural radiation includes cosmic rays and primordial radionuclides such as uranium, thorium, actinium, and potassium, which have long half-lives<sup>4</sup>. Artificial radiation from human activities, such as nuclear testing and accidents, includes radionuclides such as Cs-137 and Sr-90<sup>5</sup>. These substances can enter the body via inhalation, ingestion, and dermal absorption<sup>6</sup>. Contaminated water can introduce internal radiation sources, leading to DNA and RNA damage and potentially causing cancer and genetic mutations<sup>7</sup>. Industrial activities also introduce heavy metals like lead, chromium, arsenic, and cadmium into the body through contaminated water, air, and food, accumulating in tissues and bones<sup>8</sup>.

Underground water that interacts with geological formations and contaminants can have higher concentrations of radioactive materials than surface water<sup>9</sup>. The global demand for bottled water has increased by 31% due to economic and safety reasons<sup>10</sup>. This study evaluated the specific activities of radioactive isotopes and heavy metal levels in bottled water and compared them with WHO health standards and findings from other countries.

Applying ML in the mineral water context involves innovative approaches to improve water quality, treatment processes, and resource management. ML algorithms can analyze data from various sensors for real-time monitoring of mineral water quality, which is essential for identifying contaminants and ensuring compliance with safety regulations<sup>11</sup>. ML models can enhance water treatment efficiency by forecasting optimal operational parameters, thereby leading to improved impurity removal and resource efficiency. In addition, ML can predict the maintenance needs of water treatment systems, thereby reducing downtime and ensuring continuous functionality<sup>11</sup>. In mineral processing, ML models can identify and assess the mineral composition of water, which is crucial for achieving the desired mineral balance in bottled water. These applications highlight ML's transformative potential in the mineral water industry, particularly in terms of quality assurance, process optimization, and sustainable water management<sup>12</sup>.

Tackling a regression problem involves predicting a continuous outcome variable based on one or more predictor variables, which is common in fields such as finance, healthcare, and engineering. Accurate predictions

<sup>1</sup>Department of Physics, Faculty of Sciences, Arak University, Arak 38156-8-8349, Iran. <sup>2</sup>Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran 4117-13114, Iran. ✉email: H-Sadeghi@araku.ac.ir

require a structured approach, starting with an understanding of the data and relationships between variables through exploratory data analysis (EDA)<sup>13</sup>. Complex models such as polynomial regression, ridge regression, and lasso regression can capture nonlinear relationships and handle multicollinearity. Model evaluation and validation, such as cross-validation, ensures that the model generalizes well to unseen data, which enhances reliability<sup>14</sup>.

The goal of this study is to apply sophisticated ML models to improve prediction precision in future studies. The study also proposes recommendations for regulating radioactive isotopes and heavy metals in potable water and advocates for longitudinal studies to assess the health effects of consuming mineral water with low concentrations of radionuclides and heavy metals.

The paper is organized as follows: section “Materials and methods” details the materials, methods, and formulations for measuring radioactive isotope activity levels and reviews various ML models to identify the best regression model for predicting cancer risk. Section “Results and discussion” compares results from mineral water samples in Arak City with those predicted by ML models, discussing the most effective methods and cancer risk assessment. Finally, section “Summary and conclusion” summarizes the research findings.

## Materials and methods

In this study, a random sampling technique was employed to assess the activity levels of the radioactive isotopes Ra-226, Cs-137, K-40, and Th-232 in water sourced from the consumable mineral water of Arak City. The investigation utilized commercially available mineral waters from the region. Initially, 15 samples of consumable mineral water, branded as Surprise, Miwa, Aquafina, Versailles, Deserni, Zamzam, Akarso, Vata, Damavand, Elham, Prolife, Kalis, Aqiq, Gohar, and Alis, were collected from local supermarkets in Arak. Each sample was 1.5 L of water to satisfy the measurement criteria. The samples were designated with coded identifiers WFR1-WFR15, and 800 cc samples were transferred into Marinelli beaker containers thoroughly cleaned with distilled water and alcohol. These Marinelli containers were then sealed with aquarium adhesive to eliminate any potential air exchange with the external environment and to prevent the escape of radon gas. To achieve stable equilibrium among the decay products, eight half-lives of the radon gas must elapse, which corresponds to a duration of approximately 50 days, given that the half-life of radon is 7 days. Consequently, mineral water sampling was scheduled to occur approximately 50 days after sealing<sup>15</sup>.

To investigate the heavy-metal concentration in mineral water samples using the inductively coupled plasma technique, 10 cc of each sample was transferred into Falcon containers located in the central laboratory of Arak University. The samples were subsequently pumped into an atomizer, from which they were injected into the ICP-AES device. The plasma within the device reaches a temperature of 4000 °C, which facilitates the dissociation of chemical bonds and results in light emission at various wavelengths. These wavelengths are then used to quantify the heavy metal contents in the samples. To mitigate potential chemical interferences within the apparatus, three appropriate emission lines were selected for each element, and each analysis was conducted in triplicate. Table 1 presents the characteristics of 15 consumable mineral water samples sourced from various cities, each exhibiting distinct compositions, collected from supermarkets in Arak City.

## Sample preparation

The gamma-ray spectrometry apparatus used in the nuclear laboratory at Arak University is designed to measure the activity of radioactive nuclei with high-energy resolution. This system employs a Semiconductor and coaxial HPGe p-type detector model 30,195 BSIGCD, which boasts a relative efficiency of 30% and is equipped with 4096 analyzer channels. The energy resolution of the Co-60 gamma line, characterized by energies of 1332.520 keV and operating at a voltage of 3000 V, was recorded at 1.95 keV. The energy and efficiency calibrations in gamma spectroscopy are conducted using a Cs-137 standard source with known specific activity, and the spectrometry process is facilitated by lsrmb SI software. The signals produced in the detector are transmitted to the MCA8192 compact system via a pre-amplifier, where they are recorded in a sped format. Subsequently, WINHEX and MATLAB software were employed to convert these files into the CHN format. Following this conversion, the specific activities of radionuclides in environmental samples were assessed using GammaVision32 (EG&G Ortec). To determine background radiation and ensure that the spectra are analyzed under consistent spatial and temporal conditions, spectroscopy is performed in empty Marinelli containers, and the resulting data is subtracted from the original spectrum<sup>16</sup>.

In the central laboratory of Arak University, there exists an Inductively Coupled Plasma Atomic Emission Spectrometry (ICP-AES) instrument, specifically the 9100 Quant Plasma model, manufactured in Germany. This analytical technique is used for the detection of chemical elements in various samples. The device operates using inductively coupled plasma to generate excited atoms and ions, which subsequently emit electromagnetic radiation at characteristic wavelengths associated with specific elements. Plasma is a high-temperature source of ionized gas, typically argon. The device is maintained in the temperature range of 6000 to 10,000 K using inductive pairs of electric coils and operates at MHz frequencies. The absolute efficiency of the detector is determined by Eq. (1) based on the recorded gamma-ray spectrum.

$$\epsilon (\%) = \frac{N_i}{\text{Act} \times P_n(E_i) \times t} \times 100, \quad (1)$$

where  $\epsilon$  represents the absolute efficiency of the detector,  $N_i$  denotes the net count of the sub-peak corresponding to the energy  $E_i$ , and Act refers to the specific radioactivity of the standard source measured in Bq/Kg. Additionally,  $P_n(E_i)$  indicates the probability of photon emission at energy  $E_i$ , and  $t$  indicates the duration of the spectrometry measurements, expressed in seconds.

Sample no.	Brand name	Factory location (in Iran)	Mineral combinations
wfr1	Surprise	Alborz, Baharestan Province	Fluoride, calcium, potassium, sulfate, magnesium, nitrate, and nitrite
wfr2	Miwa	Amol and Pleur springs	Calcium, magnesium, fluoride, chloride, nitrite, calcium carbonate, and soluble solids
wfr3	Aquafina	Tehran, km 10 special road	Calcium, magnesium, sodium, potassium, and sulfate
wfr4	Versa	Amol, Haraz Road	Calcium, sodium, magnesium, nitrite, nitrate, sulfate, and fluoride
wfr5	Desserny	Tehran, Azadegan Highway	Sodium, potassium, iron, calcium, magnesium, and sulfates
wfr6	Zamzam	Lorestan, Azna	Chloride, magnesium, iron, sodium, and calcium
wfr7	Akarso	Astana, Anusha	Calcium, magnesium, sodium, fluoride, solid solutions, calcium carbonate
wfr8	Vata	Ardabil, the Sablan mountain	Calcium, magnesium, fluoride, chloride, nitrate, nitrite, and sodium
wfr9	Damavand	Damavand Mountain, Tehran, Iran	Calcium, magnesium, and sulfate
wfr10	Elham	Hamedan, Faminin	Calcium, magnesium, nitrite, nitrate, and sodium
wfr11	Prolife	Haraz Road	Calcium, sodium, bicarbonate, chloride, and nitrite
wfr12	Kallis	Lorestan, Cheshme Mahihole	Calcium, magnesium, sodium, sulfate, nitrate, and fluoride
wfr13	Aghigh	Arak, Safidkhani foothills	Calcium, magnesium, nitrite, sodium, bicarbonate, and nitrate
wfr14	Gahar	Lorestan–Astrankoh range	Calcium carbonate, soluble solids, chloride, and nitrate
wfr15	Alis	Khorasan Razavi	Calcium, magnesium, sodium, calcium carbonate, and chloride

**Table 1.** Brief description of mineral water in Iran, consumable at Arak City.

### Special activities of radioactive nuclei

The specific activity of samples is assessed using gamma rays from various isotopes. For the gamma lines of Pb-214, which have an energy of 351.93 keV, and Bi-214, which have an energy of 609.31 keV, the specific activity is determined alongside those of Th-232. The gamma line from Ac-228, which exhibited an energy of 911.21 keV with an intensity of 28%, and another line at 338.32 keV with an emission percentage of 11%, was also analyzed. Additionally, the specific activity of K-40 was evaluated using its gamma line at 1460.70 keV, whereas the gamma line of Cs-137, with an energy of 661.66 keV, was employed for further assessment. The specific radioactivity of these radioactive nuclei within the samples is given by

$$A = \frac{\text{NetArea}}{\epsilon (\%) \times \text{BR} (\%) \times t \times m} \times 100, \quad (2)$$

where A denotes the specific activity of the sample expressed in Bq/kg. The term “Net Area” refers to the net area beneath the peak, while  $\epsilon$  signifies the absolute efficiency of the detector. Additionally, BR represents the branching ratio expressed as a percentage, t indicates the sampling duration of the sample measured in seconds, and m corresponds to the mass of the sample in kilograms<sup>17</sup>.

### Annual effective dose

The annual effective dose derived from the ingestion of natural and artificial radionuclides in drinking water is as follows:

$$\text{AED} = \sum A_i \times \text{DCF}_i \times \text{Cr}. \quad (3)$$

The annual effective dose (AED) is expressed in sieverts per year and is calculated using the specific radioactivity of isotopes Ra-226, Th-232, K-40, and Cs-137 measured in units per liter (Bq/L). The dose conversion factor ( $\text{DCF}_i$ ) is defined in Sieverts per liter (Sv/Bq), and Cr represents the annual drinking water consumption for infants, children, and adults (250, 350, and 730 L, respectively)<sup>18</sup>. Table 2 lists the dose conversion values of the core radioactive substances.

### Potential cancer risk (ELCR)

To evaluate the potential cancer risk associated with drinking water consumption throughout an individual's lifetime, various methodologies

DCF <sub>i</sub> (Sv/Bq)			
<sup>226</sup> Ra	<sup>232</sup> Th	<sup>40</sup> K	<sup>137</sup> Cs
2.8 × 10 <sup>-7</sup>	6.9 × 10 <sup>-7</sup>	6.2 × 10 <sup>-9</sup>	1.3 × 10 <sup>-8</sup>

**Table 2.** Values of specific radiation conversion factor to dose.

Radioactive nuclei	h <sub>ing</sub> (mSv/Bq)	Guidance level (Bq/l)
<sup>226</sup> Ra	2.1 × 10 <sup>-4</sup>	0.5
<sup>232</sup> Th	2.3 × 10 <sup>-4</sup>	1.0
<sup>40</sup> K	6.2 × 10 <sup>-6</sup>	22.1
<sup>137</sup> Cs	1.3 × 10 <sup>-5</sup>	10.0

**Table 3.** Guidance levels for radioactive nuclei in young adults’ drinking water.

$$\text{ELCR} = R_F \times D_W \times F_{AR}. \tag{4}$$

The ELCR represents the lifetime cancer risk, and  $D_w$  denotes the annual effective dose for the specified age group, which is measured in Sieverts annually. The  $F_{AR}$  indicates the duration of the target age range in years, and the  $R_F$  which is quantified as  $7.3 \times 10^{-2}$ , corresponds to the risk associated with one per Sv<sup>19</sup>.

**Radium equivalent activity (Raeq)**

The total radioactivity can be determined using Eq. (5), which expresses it concerning the radium activity (Raeq) as follows<sup>20</sup>

$$R_{aeq} = A_{Ra} + 1.43 A_{Th} + 0.077A_K. \tag{5}$$

The specific activities of Ra-226, Th-232, and K-40 (denoted as  $A_{Ra}$ ,  $A_{Th}$ , and  $A_K$ , respectively) were measured in Bq/l. The internal risk indicators ( $H_{in}$ ) and external risk indicators ( $H_{ex}$ ) were used to assess the risk of radiation exposure associated with specific isotopes within the radon gas decay series. The values of  $H_{ex}$  and  $H_{in}$  can be derived from Eqs. (6) and (7)<sup>21</sup> as follows:

$$H_{ex} = \frac{A_{Ra}}{370} + \frac{A_{Th}}{259} + \frac{A_K}{4810} \leq 1, \tag{6}$$

$$H_{in} = \frac{A_{Ra}}{185} + \frac{A_{Th}}{259} + \frac{A_K}{4810} \leq 1. \tag{7}$$

**Safe drinking water**

To determine the surface area of the core guide, we used<sup>22</sup>:

$$GL = \frac{IDC}{(q \times h_{ing})}, \tag{8}$$

where GL represents the guide level of the radioactive core in drinking water, measured in becquerels per liter. IDC denotes the individual dose standard, established at 0.1 millisieverts per year. The variable q indicates the annual water consumption of adults (730 L per year, while  $h_{ing}$  refers to the dose conversion factor expressed in millisieverts. This factor was derived from the values associated with Becquerel (see Table 3).

To ensure the safety of drinking water, we implemented various measures.

$$SFW = \sum \frac{C_i}{GL_i} \leq 1. \tag{9}$$

The specific activity of the radioactive nuclei for the i-th radionuclide in drinking water is denoted  $C_i$ , while  $GL_i$  represents the guideline level for the i-th radionuclide, as derived from Eq. (8). It is established that specific drinking water activities should not exceed a value of 1.

**ML methods**

A suitable regression model is essential for an effective analysis. Linear regression is one of the most straightforward and commonly employed techniques, primarily because of its simplicity in application. Furthermore, performance indicators such as Mean Squared Error (MSE) and R-squared, serve to measure prediction precision. Ultimately, it is crucial to continuously enhance the model by adjusting hyperparameters and integrating domain-specific insights, which improves performance. The refinement process is critical for developing accurate and reliable models. A systematic methodology is essential when tackling a regression problem to achieve precise and reliable outcomes. The main aim of this study was to predict AEDs and assess ELCR. The following outlines the key steps of this research:

- (a) **Data Collection:** This step acquires all pertinent data required for analysis, ensuring its relevance and comprehensiveness. The data presented in sections “Sample preparation” to “Safe drinking water” are primarily discussed in the results and discussion sections.
- (b) **Data Preprocessing:** Data preprocessing is a fundamental step in ML because it helps improve the quality of data and prepare them for modeling. In our dataset, there were no outliers or missing values. We used the one-hot encoding method to convert categorical mineral water type and age group data into numerical data.
- (c) **Data Splitting:** The dataset was divided into a 70:30 ratio, with 70% allocated for training to learn and 30% for testing to evaluate the performance of the model on previously unseen data.
- (d) **Model Selection:** Various regression models, including ridge regression, Decision Tree (DT) Regression, and Random Forest (RF) Regression, were evaluated to determine the most appropriate option. Below, each model is presented with a concise description:

Ridge Regression: Ridge regression improves linear regression by addressing multicollinearity by adding a regularization term to the ordinary least squares objective function. The L2 penalty reduces model complexity by shrinking the coefficients toward zero. Ridge regression is particularly useful for handling multicollinearity, which occurs when independent variables are highly correlated<sup>23</sup>.

DT Regression: DT regression employs a top-down, greedy-layer approach. The proposed method progressively divides a dataset into increasingly smaller subsets while simultaneously constructing an associated decision tree. The end product is a tree comprising decision and leaf nodes<sup>24</sup>.

RF Regression: RF Regression is an additive model that predicts outcomes by aggregating decisions from multiple base models. Each base model is a DT, and the final output of the RF model is the combined result of the DT. This method of using several models to enhance predictive accuracy is referred to as model ensemble<sup>24</sup>.

- (e) **Model Training and Hyperparameter Tuning:** The grid search method is used as a strategy to optimize the hyperparameters of ML models to enhance performance. This approach specifies a range of potential values for each hyperparameter, followed by an exhaustive evaluation of all possible combinations to identify the most effective configuration for improving model performance. Table 4 presents the hyperparameter configurations for three different models: Ridge, DT, and RF. These configurations were determined using a grid search technique, thereby outlining the ideal hyperparameters for each model.
- (f) **Model Evaluation:** The efficacy of the model was evaluated using various metrics, including the mean absolute error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R<sup>2</sup>). The mean absolute error (MAE) quantifies the disparity between the actual and forecasted values by computing the average of the absolute differences throughout the dataset<sup>25</sup>.

$$MAE = \frac{\sum_i |y_i - \hat{y}|}{n} \tag{10}$$

The RMSE is derived by computing the square root of the mean square error (MSE). The MSE quantifies the disparity between the actual and predicted values by squaring the mean of the differences throughout the dataset<sup>25</sup>.

$$RMSE = \sqrt{\frac{1}{n} \sum_i (y_i - \hat{y})^2} \tag{11}$$

Coefficient of Determination (R<sup>2</sup>): R<sup>2</sup> denotes the coefficient of determination, which serves as an indicator of the degree to which the observed values align with the original values. The coefficient ranges from 0 to 1 and can be interpreted as a percentage. A higher R<sup>2</sup> value indicates a superior model performance<sup>25</sup>.

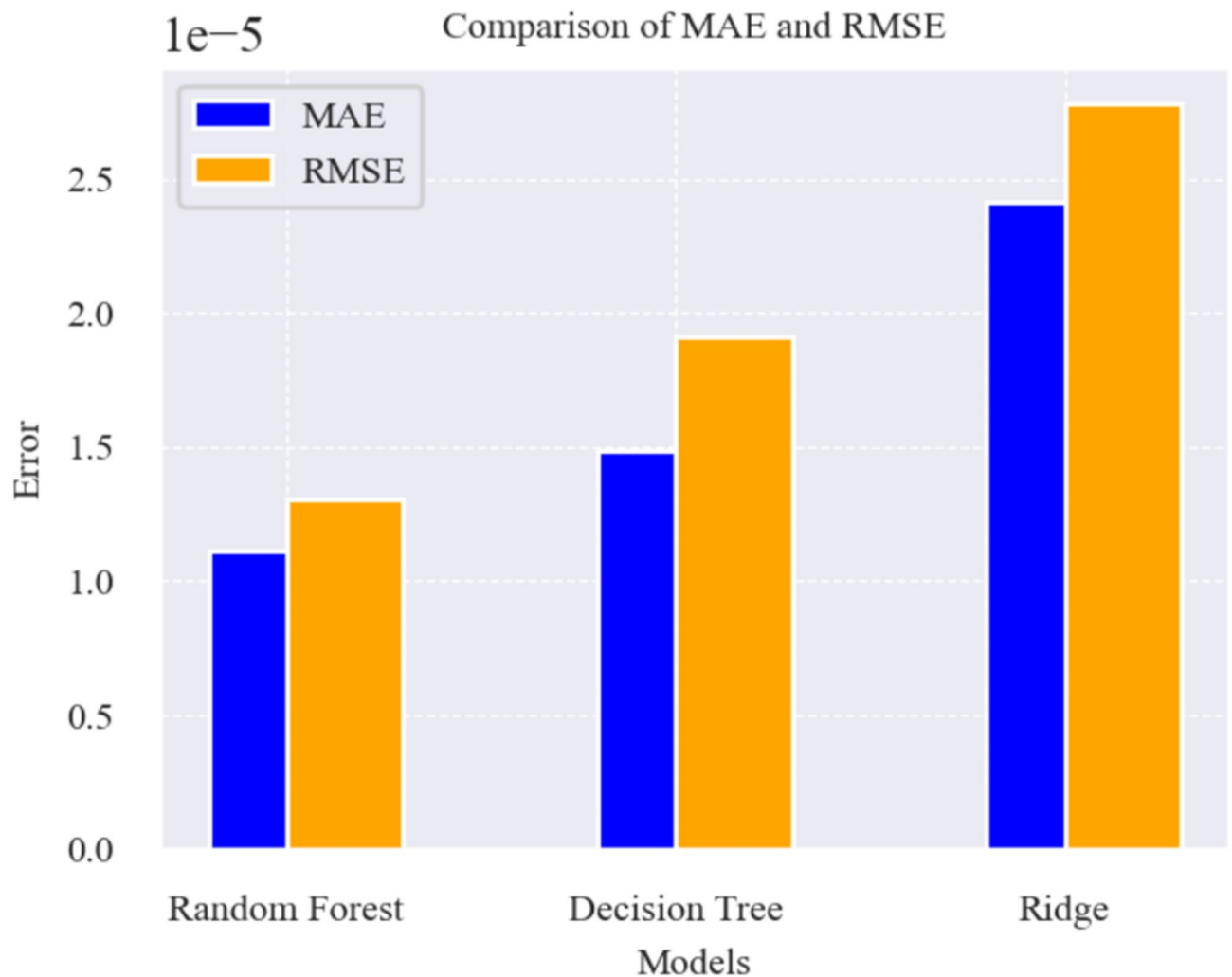
$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y})^2}{\sum_i (y_i - \bar{y})^2} \tag{12}$$

(g) **Selecting the Best Model and Result Interpretation:**

Figure 1 presents a comparative analysis of the error metrics: MAE and RMSE across three ML models: RF, Decision Tree, and Ridge models. These metrics are critical for assessing the accuracy of regression models by

Model name	alpha			
Ridge regressor	0.1			
Model name	n_estimators	max_depth	min_samples_split	min_samples_leaf
DT regressor	–	10	2	2
RF regressor	50	10	2	1

**Table 4.** The hyperparameter configurations for three different models: Ridge, DT, and RF.



**Fig. 1.** A comparative examination of two error metrics, specifically MAE and RMSE error metrics were compared across three ML models: RF, Decision Tree, and ridge regression models.

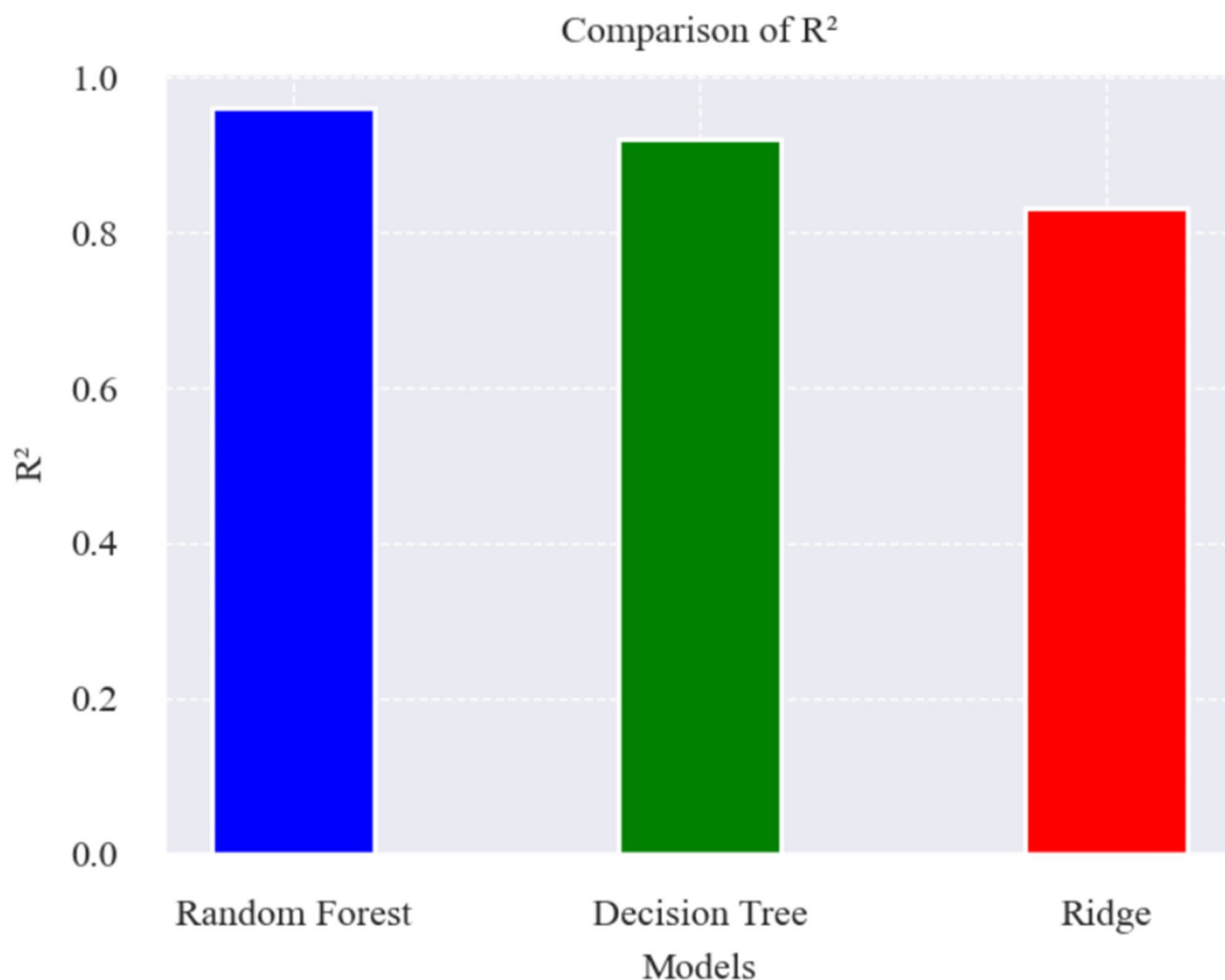
quantifying the discrepancies between actual and predicted values. The MAE is represented by blue bars, and it quantifies the average absolute deviation between the actual and predicted values, regardless of error signs. A lower MAE typically indicates a more precise model. In this analysis, the RF model exhibits the lowest MAE, indicating superior accuracy compared to the other models. In contrast, the Ridge model obtained the highest MAE, indicating a greater degree of error in its predictions. The MAE of the DT model was marginally higher than that of the RF model, reflecting its moderate performance in this context.

In contrast, the RMSE is illustrated with yellow bars and evaluates errors by assigning more penalties to larger discrepancies, which involves squaring the errors. This metric is particularly sensitive to larger errors than the MAE. The RF demonstrated the lowest RMSE, indicating high accuracy and reduced occurrence of significant errors. The ridge model, which had the highest RMSE value, tended to have larger prediction errors than the other models. Although the DT demonstrated a lower RMSE than the ridge, it still fell short of the accuracy exhibited by RF. Overall, the chart indicates that the RF outperforms both the MAE and RMSE, highlighting its effectiveness in minimizing prediction errors.

The  $R^2$  indices of the  $R^2$  index between three different ML models are displayed in Fig. 2. The RF, DT, and Ridge models were evaluated. The  $R^2$  index, also known as the “coefficient of determination,” is a crucial indicator used to assess the accuracy of regression models. This index showcases the amount of variance in the dependent variables explained by independent variables. The closer the  $R^2$  value is to 1, the better the model’s explanation of the data variance. In the graph, the RF model, represented in blue, exhibited the highest  $R^2$  value. This suggests that this model is the most accurate in prediction. RF is a complex model that combines multiple decision trees, which is typically employed in scenarios with complex data. RF can be used to identify correlations between features.

The DT model (green curve) is less accurate than the RF model. This can be attributed to the simplicity of the DT model, as it consists of a single tree rather than a combination of trees, as in the RF. Despite being less accurate than the RF, the DT model seems to outperform the Ridge model (red), which ranks last among





**Fig. 2.** Comparative analysis of the  $R^2$  index across three distinct ML models: RF, DT, and Ridge.

the three models. This comparison reveals that the more intricate RF model performs better on our data and produces higher  $R^2$  values.

After evaluating the performance of several regression models based on various metrics, we found that the RF Regression model demonstrated the best results.

To assess the robustness of the results obtained from the proposed RF model, we evaluated the model stability using two methods. First, we applied small noise (mean 0, standard deviation 0.01) to the test data and compared the original  $R^2$  with the perturbed  $R^2$ . Both values remained consistent at 0.9, demonstrating the model's robustness and stability against minor input variations. These results suggest that the model can reliably maintain its predictive power, which is crucial for real-world applications in which data often contain small amounts of noise. Second, the bootstrap method was used to estimate the stability of the proposed model. Given the limited dataset size, Cross-Validation (CV) often produces unstable and high-variance results because the small subsets in each fold provide insufficient data for robust training and testing. This limitation can lead to inaccurate model performance estimation. To address this issue, the bootstrap method was employed. The proposed method repeatedly samples with replacement from the entire dataset, thereby allowing for optimal use of limited data. In this study, the bootstrap procedure was repeated 100 times, enabling the calculation of the mean and standard deviation of the performance metric  $R^2$ . The results demonstrated a mean  $R^2$  of 0.89 with a standard deviation of 0.08, indicating the model's stability. Therefore, due to its ability to deliver a more consistent and trustworthy evaluation, the bootstrap method was preferred over CV in this case.

Generally, the RF model demonstrated superior accuracy, robustness, and generalizability. As a result, the random forest model was chosen to make the final predictions. By leveraging its ability to capture complex patterns and relationships in the data, we can anticipate precise predictions for AED. Figure 3 summarizes the step-by-step process of the proposed ML regression model.

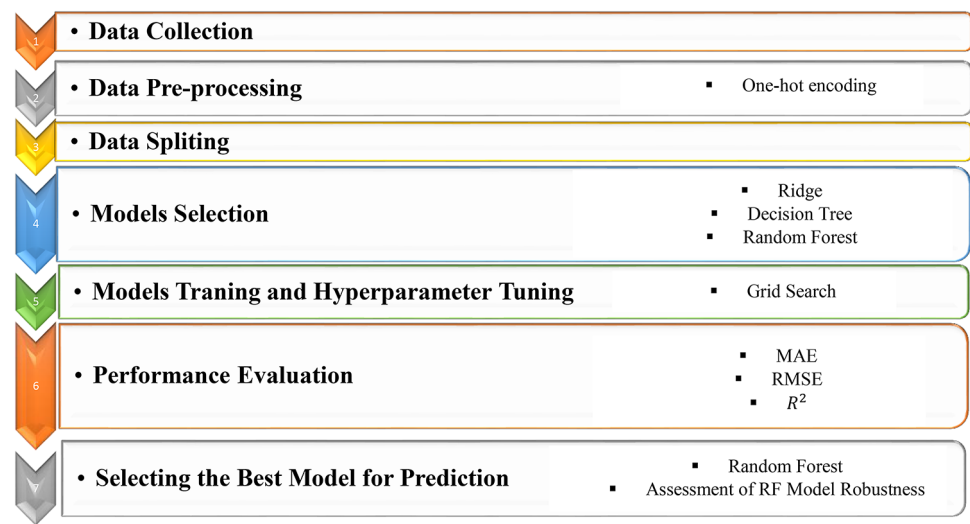


Fig. 3. Steps to construct the machine learning regression model.

Sample no.	Net area	$\epsilon$	BR	T	m	$^{226}\text{Ra}$
wfr1	0	0.015127	0.46	86,400	0.8	0
wfr2	0	0.015127	0.46	86,400	0.8	0
wfr3	0	0.015127	0.46	86,400	0.8	0
wfr4	0	0.015127	0.46	86,400	0.8	0
wfr5	0	0.015127	0.46	86,400	0.8	0
wfr6	0	0.015127	0.46	86,400	0.8	0
wfr7	0	0.015127	0.46	86,400	0.8	0
wfr8	0	0.015127	0.46	86,400	0.8	0
wfr9	0	0.015127	0.46	86,400	0.8	0
wfr10	0	0.015127	0.46	86,400	0.8	0
wfr11	0	0.015127	0.46	86,400	0.8	0
wfr12	0	0.015127	0.46	86,400	0.8	0
wfr13	0	0.015127	0.46	86,400	0.8	0
wfr14	0	0.015127	0.46	86,400	0.8	0
wfr15	0	0.015127	0.46	86,400	0.8	0

Table 5. Summary of the special activity results of  $^{226}\text{Ra}$  related to all samples according to (Bq/l).

Results and discussion

The specific activities of the radioactive nuclei across all samples are presented in Tables 5, 6, 7 and 8, and Fig. 4 to facilitate the comparative analysis of the specific activity results. The annual dose received, expressed in ( $\mu\text{Sv/y}$ ) due to water consumption, is detailed for the three distinct age groups in Table 9. Figure 5 visual comparison of the dose results. Additionally, data concerning cancer risk associated with water consumption among the same three age groups are presented in Table 10, with Fig. 6 to enhance the comparative evaluation of cancer-related outcomes. Furthermore, Table 11 presents results related to radium equivalent activity, internal and external risk indices, and assessments of drinking water safety. A comparative analysis of the findings from this study and those from other countries is presented in Table 12, complemented by a comparison chart in Fig. 7.

The specific activities of  $^{226}\text{Ra}$  are detailed in Table 5. By applying Eq. (2), it is established that the total sampling duration for all specimens spans one day and night, totaling 86,400 s. The cumulative mass of the samples was 800 cc, equivalent to 0.8 kg. The branching ratio for  $^{226}\text{Ra}$  is estimated at 0.46, based on decay software analysis of  $^{214}\text{Bi}$ , which has an energy of 609 keV. The calculated absolute efficiency of  $^{226}\text{Ra}$  is 0.015127. The net level for all samples was zero. Because the net level is a pivotal element for assessing specific activities, we conclude that the specific activity level of  $^{226}\text{Ra}$  across all samples is also zero.

The specific activities of  $^{232}\text{Th}$  are presented. Using Eq. (2), we determined that the total sampling duration for all samples was one day, equivalent to 86400 s, with a total mass of 800 cc or 0.8 kg for each sample, as shown in Table 6. The branching ratio of  $^{232}\text{Th}$ , calculated using Decay software for  $^{228}\text{Ac}$  at an energy level of 911 keV, was estimated to be 0.28. The absolute yield for  $^{232}\text{Th}$  is recorded as 0.011389. Among the samples, sample 7 exhibited the highest net activity (165), while sample 14 exhibited the lowest net activity ((0). Furthermore, the



Sample no.	Net area	$\epsilon$	BR	T	m	$^{232}\text{Th}$
wfr1	61	0.011389	0.28	86,400	0.8	0.276
wfr2	39	0.011389	0.28	86,400	0.8	0.176
wfr3	53	0.011389	0.28	86,400	0.8	0.240
wfr4	66	0.011389	0.28	86,400	0.8	0.299
wfr5	108	0.011389	0.28	86,400	0.8	0.489
wfr6	117	0.011389	0.28	86,400	0.8	0.530
wfr7	165	0.011389	0.28	86,400	0.8	0.748
wfr8	55	0.011389	0.28	86,400	0.8	0.249
wfr9	71	0.011389	0.28	86,400	0.8	0.322
wfr10	40	0.011389	0.28	86,400	0.8	0.181
wfr11	40	0.011389	0.28	86,400	0.8	0.181
wfr12	99	0.011389	0.28	86,400	0.8	0.449
wfr13	78	0.011389	0.28	86,400	0.8	0.353
wfr14	0	0.011389	0.28	86,400	0.8	0.000
wfr15	37	0.011389	0.28	86,400	0.8	0.167

**Table 6.** Summary of report special activity results for  $^{232}\text{Th}$  related to all samples according to (Bq/l).

Sample no.	Net area	$\epsilon$	BR	T	m	$^{40}\text{K}$
wfr1	309	0.008009	0.1	86,400	0.8	5.582
wfr2	98	0.008009	0.1	86,400	0.8	1.770
wfr3	4	0.008009	0.1	86,400	0.8	0.072
wfr4	19	0.008009	0.1	86,400	0.8	0.343
wfr5	135	0.008009	0.1	86,400	0.8	2.438
wfr6	84	0.008009	0.1	86,400	0.8	1.517
wfr7	9	0.008009	0.1	86,400	0.8	0.162
wfr8	0	0.008009	0.1	86,400	0.8	0.000
wfr9	413	0.008009	0.1	86,400	0.8	7.460
wfr10	387	0.008009	0.1	86,400	0.8	6.991
wfr11	148	0.008009	0.1	86,400	0.8	2.673
wfr12	47	0.008009	0.1	86,400	0.8	0.849
wfr13	82	0.008009	0.1	86,400	0.8	1.481
wfr14	0	0.008009	0.1	86,400	0.8	0.000
wfr15	12	0.008009	0.1	86,400	0.8	0.261

**Table 7.** Summary of report special activity results for  $^{40}\text{K}$  related to all samples according to (Bq/l).

highest specific activity of  $^{232}\text{Th}$  was associated with sample 7, which was recorded at 0.748 Bq/L, whereas sample 14 had the lowest specific activity at zero.

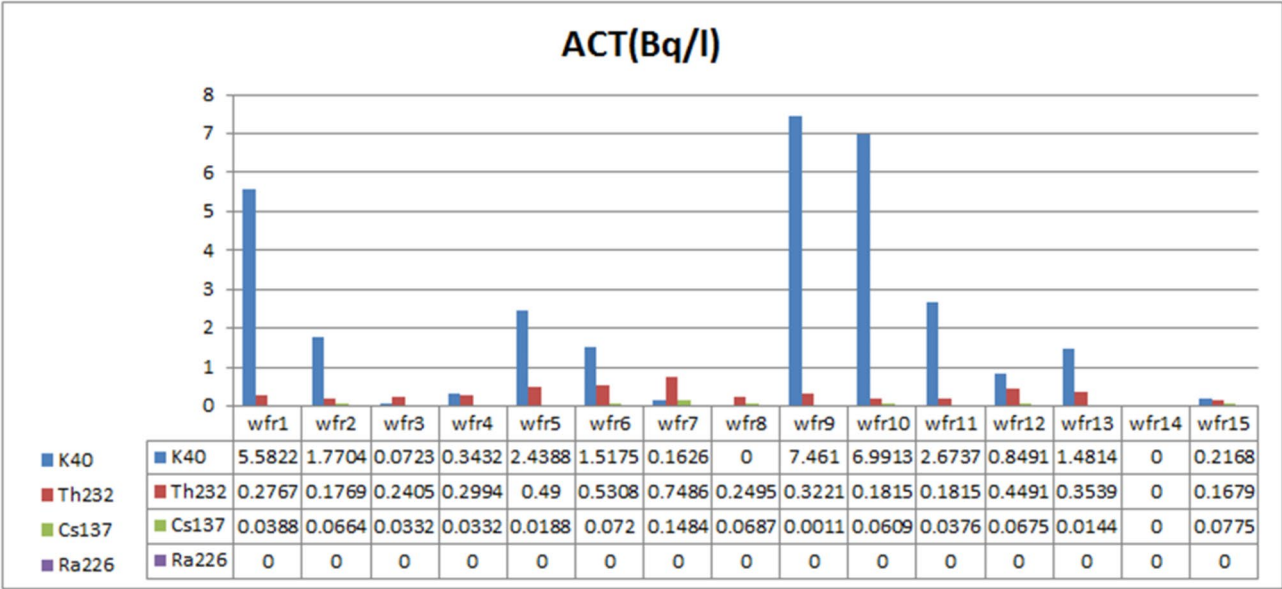
Table 7 lists the special activity values for  $^{40}\text{K}$ . Using Eq. (2), it is determined that the total sampling duration for all samples amounts to one day and night, equivalent to 86,400 s. The total mass of all samples was 800 cc (0.8 kg). The branching ratio for  $^{40}\text{K}$  was computed using decay software, yielding an energy value of 1460 keV and an estimated branching ratio of 0.1. The absolute efficiency for  $^{40}\text{K}$  was calculated as 0.008009. Among the samples, sample 9 exhibited the highest net level (413), while sample 14 exhibited the lowest net level (zero). The most significant specific activity for  $^{40}\text{K}$  was associated with sample 9, which measured 7.460 becquerels per liter, whereas sample 14 had the lowest specific activity, recorded as zero.

Table 8 lists the specific activities of  $^{137}\text{Cs}$ . Utilizing Eq. (2), it is determined that the total sampling duration for all samples is one day, equivalent to 86,400 s, with a total mass of 800 cc or 0.8 kg for each sample. The branching ratio of  $^{137}\text{Cs}$  was estimated to be 0.94 based on an energy level of 661.66 keV using decay software. The absolute efficiency for  $^{137}\text{Cs}$  is recorded at 0.0139. Among the samples, sample 7 exhibited the highest net level (134), while sample 14 exhibited the lowest net level (0). The specific activity of  $^{137}\text{Cs}$  was highest in sample 7, with a value of 0.148 becquerels per liter, whereas sample 14 had the lowest specific activity, recorded as zero.

The data presented in Fig. 4 indicate that the mineral water samples exhibited the highest levels of radioactivity associated with the elements  $^{40}\text{K}$ ,  $^{232}\text{Th}$ , and  $^{137}\text{Cs}$ . The concentration of  $^{226}\text{Ra}$  was recorded as zero across all samples, which is significant given that 98.5% of radiation-related damage is attributed to  $^{226}\text{Ra}$ . Consequently, the concentration of radium emerges as a critical parameter for mineral water factories seeking licensure from the Ministry of Health.

Sample no.	Net area	$\varepsilon$	BR	T	m	$^{137}\text{Cs}$
wfr1	35	0.0139	0.94	86,400	0.8	0.038
wfr2	60	0.0139	0.94	86,400	0.8	0.066
wfr3	30	0.0139	0.94	86,400	0.8	0.033
wfr4	30	0.0139	0.94	86,400	0.8	0.033
wfr5	17	0.0139	0.94	86,400	0.8	0.018
wfr6	65	0.0139	0.94	86,400	0.8	0.071
wfr7	134	0.0139	0.94	86,400	0.8	0.148
wfr8	62	0.0139	0.94	86,400	0.8	0.068
wfr9	1	0.0139	0.94	86,400	0.8	0.001
wfr10	55	0.0139	0.94	86,400	0.8	0.060
wfr11	34	0.0139	0.94	86,400	0.8	0.037
wfr12	61	0.0139	0.94	86,400	0.8	0.067
wfr13	13	0.0139	0.94	86,400	0.8	0.014
wfr14	0	0.0139	0.94	86,400	0.8	0.000
wfr15	70	0.0139	0.94	86,400	0.8	0.077

**Table 8.** Summary of special activity results from  $^{137}\text{Cs}$  related to all samples according to (Bq/l).



**Fig. 4.** The amounts of special activity  $^{226}\text{Ra}$ ,  $^{40}\text{K}$ ,  $^{232}\text{Th}$  and  $^{137}\text{Cs}$  in the samples.

The information provided in Table 8 reveals that the annual effective dose for infants ranged from zero (as observed in sample 14) to 0.130 microsieverts (as recorded in sample 7. Additionally, the maximum effective dose was lower than the observed dose in infants. The annual effective dose for the pediatric population ranged from zero microsieverts, noted in sample 14, to 0.182 microsieverts ( $n=7$ ). Moreover, the maximum effective dose for children, as indicated by UNSCEAR, is also comparatively lower. Lastly, the annual effective dose for the pediatric age group ranged from zero microsieverts (associated with sample 14) to 379 microsieverts (associated with sample 7. Furthermore, the peak effective dose for children remains below the UNSCEAR threshold.

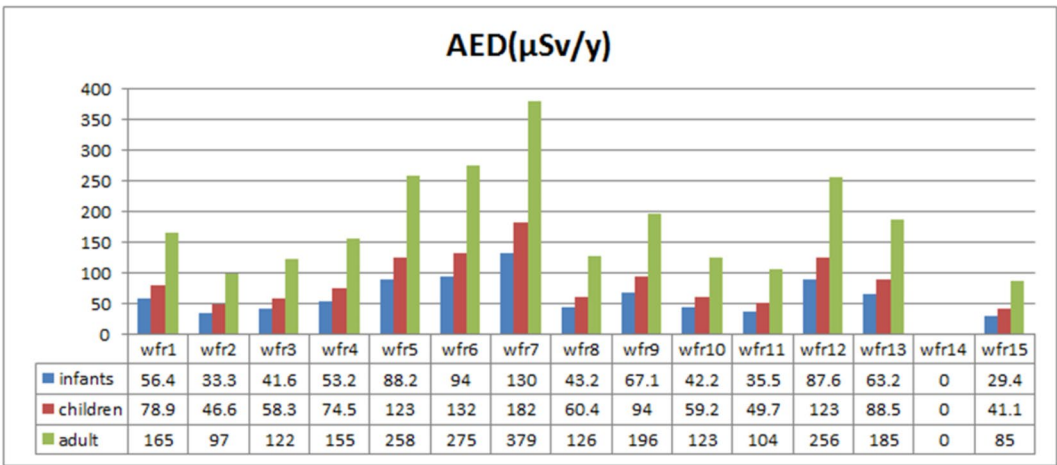
Figure 5 compares the dosages received by infants, children, and adults. The annual consumption of drinking water is 730 L for adults, 350 L for children, and 250 L for infants. Consequently, the dosage received by each age group decreased in the following order: adults, children, and infants. The highest doses for all three age groups were associated with sample 7.

The data presented in Table 10 indicate that the cancer risk factor for newborns ranges from 0 to  $71,010^{-6}$ , demonstrating variability. The cancer risk coefficient for the same age group spans from 0 to  $99,510^{-6}$ , also exhibiting variability. Additionally, the cancer risk coefficient for infants, within the range of 0–2070, is on the order of  $10^{-6}$ , highlighting its variable nature.

Figure 6 illustrates the comparative risk of cancer among the analyzed samples, including adults, infants, and children. This assessment was influenced by the critical AED variable. Notably, the effective dose received was correlated with a decreased cancer risk among the following age groups: adults, children, and infants.

-	Infants	Children	Adult
Sample no.	AED	AED	AED
wfr1	56.4	78.9	165
wfr2	33.3	46.6	97
wfr3	41.6	58.3	122
wfr4	53.2	74.5	155
wfr5	88.2	123.0	258
wfr6	94.0	132.0	275
wfr7	130.0	182.0	379
wfr8	43.2	60.4	126
wfr9	67.1	94.0	196
wfr10	42.2	59.2	123
wfr11	35.5	49.7	104
wfr12	87.6	123.0	256
wfr13	63.2	88.5	185
wfr14	0.0	0.0	0
wfr15	29.4	41.1	85
Average	57.6	80.7	168

**Table 9.** Annual dose for infants, children, and adults based on  $\mu\text{Sv/y}$ .



**Fig. 5.** Comparison of the received dose among infants, children, and adults.

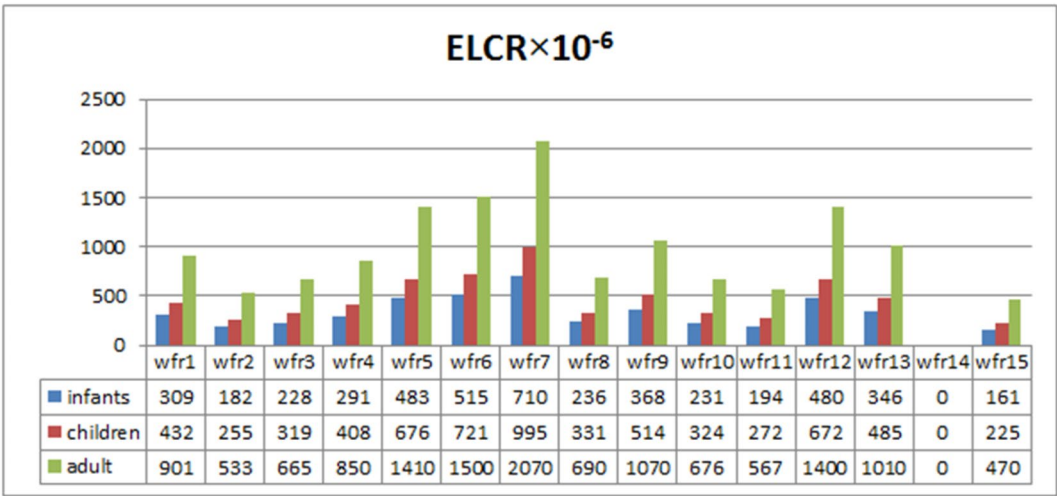
The equivalent radium activity is presented as an average of 0.61, in Table 11. The internal and external indelibilities are consistent, attributable to the uniform specific activity of radium-226, resulting in identical and low values. Furthermore, the analysis indicates that the water safety levels across all samples are below one, suggesting that there is no significant health risk associated with these findings.

To facilitate the comparison of the findings of this study, Table 12 presents the specific activity results of radioactive nuclei in bottled water from various countries. The outcomes of this investigation align well with those reported in other countries and the values established by international organizations. Consequently, it can be inferred that the radiation levels in the bottled water consumed by Arak residents do not pose any health risks. As illustrated in Fig. 7, except for Malaysia and Nigeria, the levels of radioactive nuclei activity in all other countries fell below the thresholds established by the World Health Organization. Furthermore, research on bottled water consumption in Arak City, Iran, indicates that the presence of radioactive nuclei does not represent a risk to human health.

This illustration presents the Actual versus Predicted chart for the RF model applied to the regression analysis. A chart serves as a valuable tool for juxtaposing the model's predicted values against the actual observed values, thereby facilitating the assessment of the regression model's efficacy. In this representation, points that align closely with the 45-degree line indicate that the model's predictions are nearly equivalent to the actual values. A significant number of points clustered around this line, particularly at the lower end of the X-axis, suggesting a high level of accuracy in this region. The minimal dispersion of most points on the chart indicates that the RF model effectively captured the complexities of the data and produced reliable predictions. Nevertheless, a few points deviating from the diagonal line may reflect specific characteristics of the data that warrant further

-	Infants	Children	Adult
Sample no.	ELCR	ELCR	ELCR
Rank	10 <sup>-6</sup>	10 <sup>-6</sup>	10 <sup>-6</sup>
wfr1	309	432	901
wfr2	182	255	533
wfr3	228	319	665
wfr4	291	408	850
wfr5	483	676	1410
wfr6	515	721	1500
wfr7	710	995	2070
wfr8	236	331	690
wfr9	368	514	1070
wfr10	231	324	676
wfr11	194	272	567
wfr12	480	672	1400
wfr13	346	485	1010
wfr14	0.00	0	0.00
wfr15	161	225	470
Average	316	442	922

**Table 10.** Risk factors for cancer in infants, children, and adults.



**Fig. 6.** Comparison of cancer risk among adults, infants, and children.

investigation. Overall, this chart demonstrates that the RF model has excelled in this regression task, achieving commendable accuracy in its predictions of actual values.

Figure 8 displays the actual and predicted values for the RF model used in our regression analysis. This chart is a useful tool for comparing the model's predicted values with the actual observed values, making it easier to evaluate the effectiveness of the regression model. The plot shows the actual values on the horizontal axis and the predicted values on the horizontal axis. Each point on the plot represents a specific data pair, where the actual value is compared to the predicted value.

In this representation, points that closely align with the 45-degree line indicate that the model's predictions are nearly equivalent to the actual values. Notably, a significant number of points were clustered around this line, particularly at the lower end of the X-axis, suggesting a high level of accuracy in this region. The low scatter in most parts of this image shows that the RF model can accurately predict complex representations. However, a few points deviating from the diagonal line may reflect specific data characteristics that warrant further investigation. Overall, this chart demonstrates that the RF model has excelled in this regression task, achieving commendable accuracy in its predictions of actual values.

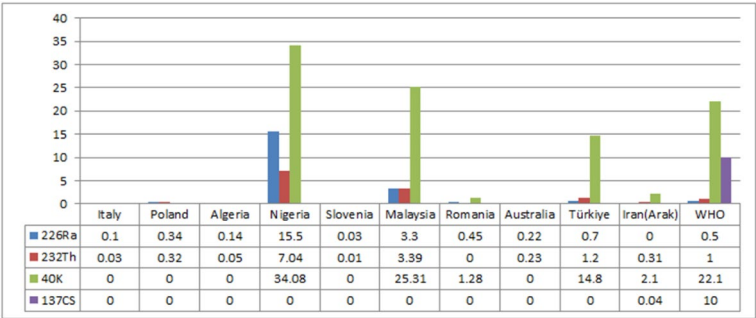
Table 13 compares the actual and predicted AECL and ECLR. The samples were divided into three groups: Infants, Children, and Adults. The table includes columns for FAR (75 for all samples), RF (0.073 for all samples), and CR (ranging from 0.00264 to 0.00401). These values are small decimals, indicating some form

Sample no.	Ra <sub>eq</sub>	H <sub>ex</sub>	H <sub>in</sub>	SFW
wfr1	0.82	0.002	0.002	0.72
wfr2	0.38	0.001	0.001	0.38
wfr3	0.34	0.001	0.001	0.41
wfr4	0.45	0.001	0.001	0.52
wfr5	0.88	0.002	0.002	0.93
wfr6	0.87	0.002	0.002	0.96
wfr7	1.08	0.002	0.002	1.28
wfr8	0.35	0.001	0.001	0.42
wfr9	1.03	0.002	0.002	0.87
wfr10	0.79	0.002	0.002	0.62
wfr11	0.46	0.001	0.001	0.42
wfr12	0.70	0.001	0.001	0.79
wfr13	0.62	0.001	0.001	0.66
wfr14	0.00	0.000	0.00	0.00
wfr15	0.25	0.001	0.001	0.29
Average	0.61	0.001	0.001	0.62

**Table 11.** Risk indicators for mineral water samples consumed in Arak City.

Country	<sup>226</sup> Ra	<sup>232</sup> Th	<sup>40</sup> K	<sup>137</sup> CS	Source
Italy	0.10	0.03	–	–	<sup>26</sup>
Poland	0.34	0.32	–	–	<sup>27</sup>
Algeria	0.14	0.05	< 0.07–2.19	–	<sup>28</sup>
Nigeria	15.50	7.04	34.08	–	<sup>29</sup>
Slovenia	0.03	0.01	–	–	<sup>30</sup>
Malaysia	3.30	3.39	25.31	–	<sup>31</sup>
Romania	0.45	–	1.28	–	<sup>32</sup>
Australia	0.22	0.23	–	–	<sup>33</sup>
Türkiye	0.70	1.20	14.80	–	<sup>34</sup>
Iraq	0.029–3.017	0.025–2.326	4.706–161.56	0.040–0.953	<sup>35</sup>
Iran	0.00	0.31	2.10	0.04	This research
WHO	0.50	1.00	22.10	10.00	<sup>36</sup>

**Table 12.** Comparison of the obtained results with those from other countries according to (Bq/l).



**Fig. 7.** Comparison of the specific activities of the radioactive cores in the mineral waters of this study with those of other countries.

of measurement or calculation. This table compares the predicted results of our prediction model with actual values across various age categories.

Figure 9 presents a summary of the feature’s importance chart. The CR feature had the highest level of importance, indicating its crucial role in the model analysis. The “Adults” feature ranked second most significant, demonstrating a considerable effect. Mineral water types 7, 6, and 5 also had notable importance although less

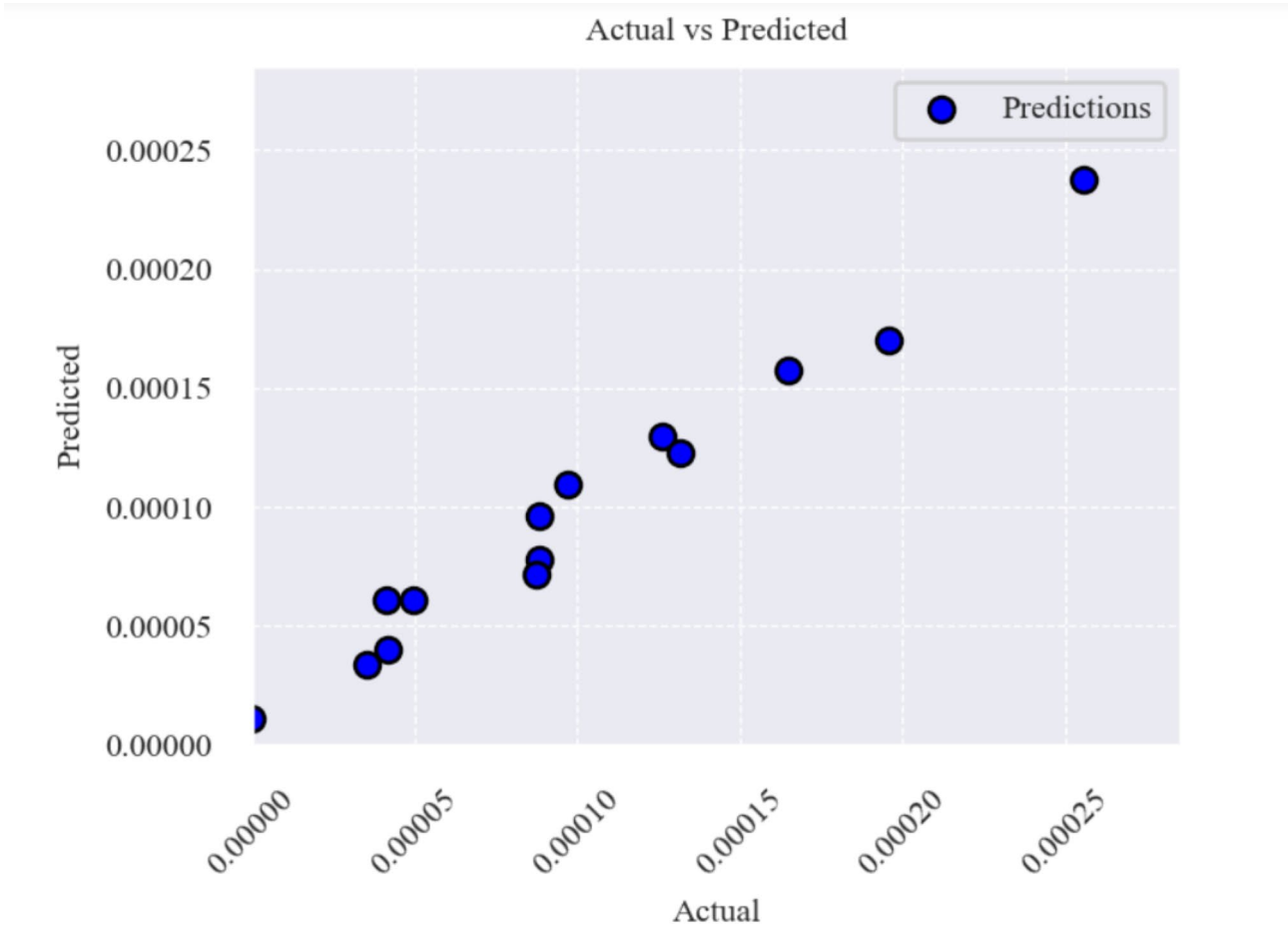
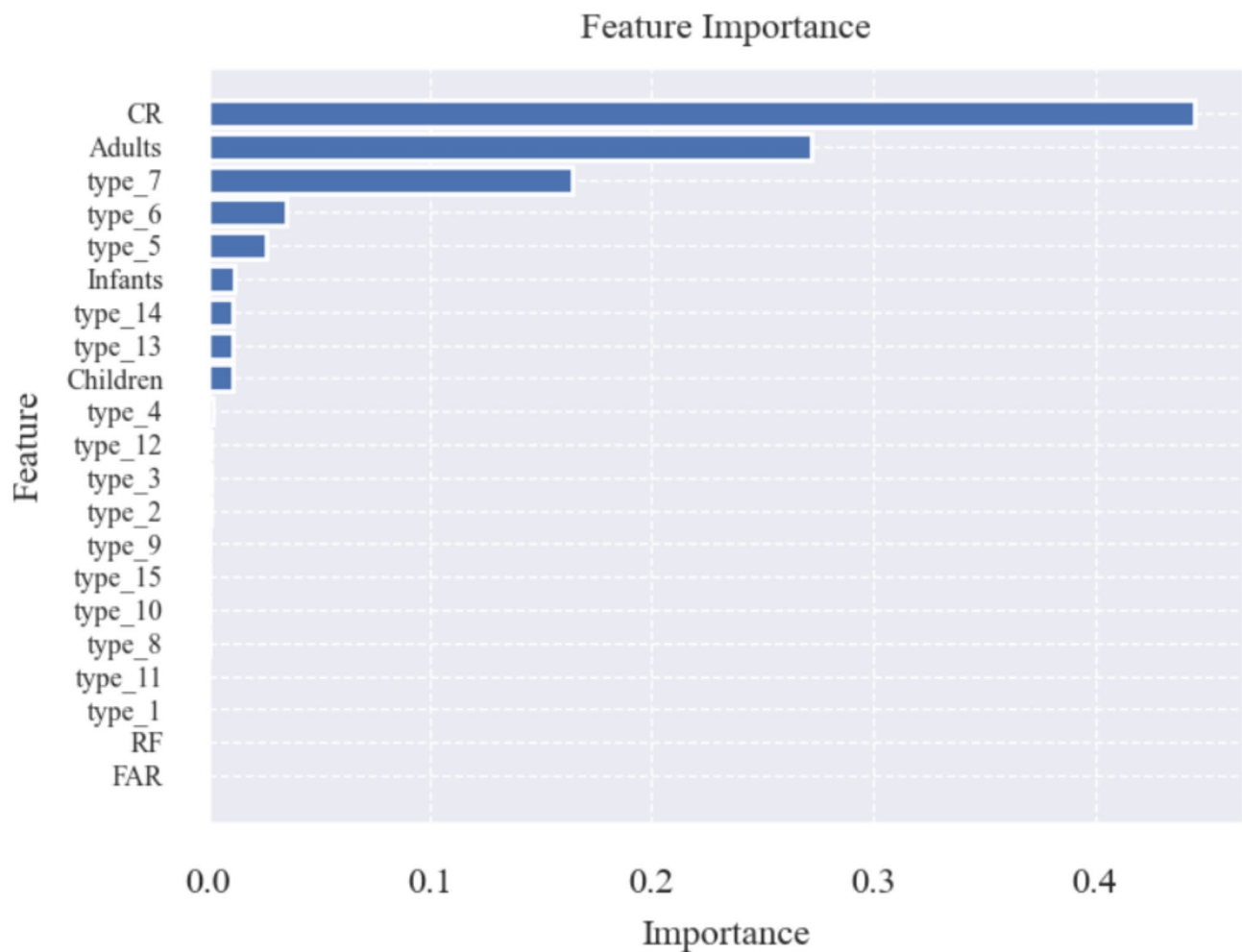


Fig. 8. Actual versus predicted values in the RF regression.

Sample no.	Age Category	FAR	RF	CR	Actual AED	Actual ELCR	Predicted AED	Predicted ELCR
wfr3	Infants	75	0.073	0.000228	0.000042	0.000228	0.000040	0.000220
wfr5	Infants	75	0.073	0.000483	0.000088	0.000483	0.000078	0.000425
wfr11	Infants	75	0.073	0.000194	0.000035	0.000194	0.000034	0.000185
wfr12	Infants	75	0.073	0.000480	0.000088	0.000480	0.000072	0.000393
wfr6	Children	75	0.073	0.009871	0.000132	0.000721	0.000123	0.000672
wfr11	Children	75	0.073	0.003726	0.000050	0.000272	0.000061	0.000335
wfr13	Children	75	0.073	0.006640	0.000089	0.000485	0.000096	0.000528
wfr14	Children	75	0.073	0.000000	0.000000	0.000000	0.000011	0.000061
wfr15	Children	75	0.073	0.003086	0.000041	0.000225	0.000061	0.000335
wfr1	Adults	75	0.073	0.000901	0.000165	0.000901	0.000158	0.000863
wfr2	Adults	75	0.073	0.000533	0.000097	0.000533	0.000110	0.000601
wfr8	Adults	75	0.073	0.000690	0.000126	0.000690	0.000130	0.000710
wfr9	Adults	75	0.073	0.001073	0.000196	0.001073	0.000170	0.000930
wfr12	Adults	75	0.073	0.001401	0.000256	0.001401	0.000238	0.001301

Table 13. Comparison of actual AECL and ECLR with predicted AECL and ECLR.





**Fig. 9.** A summary of feature importance.

than CR and Adults. On the other hand, features related to infants and children were moderately important compared to the other features. This figure highlights the features that have the greatest influence on our model analysis.

### Summary and conclusion

In all mineral water samples, special Ra-226 radiation levels were absent. The average Th-232, K-40, and Cs-137 concentrations were 0.311, 2.104, and 0.049 Bq/l, respectively, all below the WHO thresholds. The annual effective doses from bottled water consumption were 57.6  $\mu\text{Sv/y}$  for infants, 80.7  $\mu\text{Sv/y}$  for children, and 168  $\mu\text{Sv/y}$  for adults, which were significantly lower than the UNSCEAR limit of 1000  $\mu\text{Sv/y}$ . The cancer incidence coefficients were 316 for infants, 442 for children, and 922 for adults, indicating a cancer risk of  $922 \times 10^{-6}$  for a 75-year-old. Hex and Hin values ranged from 0 to 0.002, indicating no health risk. The radium equivalent activity values ranged from 0 to 1.08, aligning with global averages, with the highest level observed in the WFR7 sample. Heavy elements such as Cd, Hg, Sn, Pb, and As were detected at zero mg/L. The RF model's performance was validated by comparing actual and predicted values, demonstrating its reliability across different age groups and enhancing the study's robustness.

### Data availability

All data generated or analyzed during this study is included in this published article.

Received: 21 October 2024; Accepted: 28 January 2025

Published online: 30 January 2025

### References

1. Piñero García, F., Thomas, R., Mantero, J., Forsell-Aronsson, E. & Isaksson, M. Radiological impact of naturally occurring radionuclides in bottled water. *Food Control* **130**, 108302 (2021).
2. Dinh Chau, N. et al. *Natural Radioactivity in Groundwater—a Review* 47 (Isotopes in Environmental and Health Studies, 2011).

3. Turhan, S., Özçitak, E., Taşkın, H. & Varinlioğlu, A. Determination of natural radioactivity by gross alpha and beta measurements in groundwater samples. *Water Res.* **47**(9), 3103–3108 (2013).
4. El-Taher, A., Uosif, M. A. & Orabi, A. A. Natural radioactivity levels and radiation hazard indices in granite from Aswan to Wadi El-Allaqi southeastern desert, Egyptian. *Radiat. Prot. Dosimetry* **124**(2), 148–154 (2007).
5. Harb, S., El-Kamel, A. H., Zahran, A. M., Abbady, A. & Ahmed, F. A. Assessment of natural radioactivity in soil and water samples from Aden Governorate South of Yemen Region. *Int. J. Recent. Res. Phys. Chem. Sci. (IJRRPCS)* **1**(1), 1–7 (2014).
6. Pourimani, R. & Nemati, Z. Measurement of radionuclide concentration in some water resources in Markazi province, Iran. *Iran. J. Med. Phys.* **13**(1), 49–57 (2016).
7. Hossain, M. A. et al. Natural radioactivity levels and radiological risk assessment of surface water of wetland Tangur Haor, Sunamganj District, Bangladesh. *J. Radiation Nuclear Application* **4**(2), 117–125 (2019).
8. Massaro, E. J. *Handbook of Human Toxicology, National Health and Environmental Effects Research Laboratory* (CRC, 1997).
9. Rangela, J. D. et al. Radioactivity in bottled waters sold in Mexico. *Appl. Radiat. Isot.* **56**(6), 9. [https://doi.org/10.1016/s0969-8043\(02\)00047-7](https://doi.org/10.1016/s0969-8043(02)00047-7) (2002).
10. Ranjbar, H. & Bagheri, R. Radioactivity assessment and lifetime risk survey from bottled water in Iran. *Iran. J. Public. Health* **22**, 3. <https://doi.org/10.47176/ijpr.22.3.21383> (2022).
11. Tsipidi, I., Sartjärvi, R., Juntunen, P. & Nikolakopoulos, G. A survey on machine learning approaches in water analysis. In *Artificial Intelligence Applications and Innovations. AIAI 2024 IFIP WG 12.5 International Workshops. AIAI 2024. IFIP Advances in Information and Communication Technology, vol 715* (eds. Maglogiannis, I. et al.) (Springer, 2024). [https://doi.org/10.1007/978-3-031-63227-3\\_1](https://doi.org/10.1007/978-3-031-63227-3_1).
12. Long, T., Zhou, Z., Hancke, G., Bai, Y. & Gao, Q. A review of artificial intelligence technologies in mineral identification: classification and visualization. *J. Sens. Actuat. Netw.* **11**(3), 50. <https://doi.org/10.3390/jsan11030050> (2022).
13. Montgomery, D. C., Peck, E. A. & Vining, G. G. *Introduction to Linear Regression Analysis* 5th edn (Wiley, 2012).
14. Chatterjee, S. & Simonoff, J. S. *Handbook of Regression Analysis* (Wiley, 2013).
15. International Atomic Energy Agency. *Collection and Preparation of Bottom Sediment Sample for Analysis of Radionuclides and Trace Elements* (IAEA-TECDOC-1360, IAEA, 2003).
16. Aziz, A. In *International Atomic Energy Agency, Vienna, Methods of Low-Level Counting and Spectrometry Symposium, Berlin* 221 (1981).
17. Singh, P., Rana, N., Azam, A., Naqvi, A. & Srivastava, D. Levels of uranium in waters from some Indian cities determined by fission track analysis. *Radiat. Meas.* **26**(5), 683–687 (1996).
18. WHO. Guidelines for drinking water quality. WHO Library Cataloguing-in-Publication Data NLM Classification (2011).
19. Karabidak, S. M., Kaya, A. & Kaya, S. Cancer risk analysis of some drinking water resources in Gumushane, Turkey. *Sustainable Water Resour. Manage.* **5**, 1939–1949. <https://doi.org/10.1007/s40899-019-00338-x> (2019).
20. Mheemmed, A., Najam, L. & Khatab, A. Transfer factors of <sup>40</sup>K, <sup>226</sup>Ra, <sup>232</sup>Th from soil to different types of local vegetables, radiation hazard indices, and their annual doses. *J. Radioanal. Nucl. Chem.* **302**, 87–96. <https://doi.org/10.1007/s10967-014-3259-y> (2014).
21. Ahmed, N. K. et al. Comparative study of the natural radioactivity of some selected rocks from Egypt and Germany. *Indian J. Pure Appl. Phys.* **44**, 209–215 (2006).
22. Pourimani, R., Karimi, S. & Mohebian, M. Measurement of specific radioactivity of natural and artificial radioactive nuclei and calculation of radiological hazards related to them in drinking water of Isfahan region. *Iran. J. Radiation Saf. Meas.* **9**(1), 21–32 (2021).
23. Tran, T. K. et al. Initial coin offering prediction comparison using Ridge regression, artificial neural network, random forest regression, and hybrid ANN-Ridge. In *Mendel, Brno University of Technology* (2023).
24. Acharya, M. S., Armaan, A. & Antony, A. S. A comparison of regression models for prediction of graduate admissions. In *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)* 1–5 (IEEE, 2019).
25. Reddy, P. S. M. Decision tree regressor compared with random forest regressor for house price prediction in Mumbai. *J. Surv. Fisheries Sci.* **10**(1S), 2323–2332 (2023).
26. Desideri, D. et al. (<sup>238</sup>U, (<sup>234</sup>U, (<sup>226</sup>Ra, (<sup>210</sup>Po concentrations of bottled mineral waters in Italy and their dose contribution. *J. Environ. Radioact.* **94**(2), 86–97. <https://doi.org/10.1016/j.jenvrad.2007.01.005> (2007).
27. Kozłowska, B., Walencik, A. & Dorda, J. Natural radioactivity and dose estimation in underground water from the Sudety Mountains in Poland. *Radiat. Prot. Dosimetry* **128**(3), 331–335. <https://doi.org/10.1093/rpd/ncm380> (2008).
28. Seghour, A. & Seghour, F. Z. Radium and (<sup>40</sup>K in Algerian bottled mineral waters and consequent doses. *Radiat. Prot. Dosimetry* **133**(1), 7–50. <https://doi.org/10.1093/rpd/ncp009> (2009).
29. Ajayi, O. S. & Adesida, G. Radioactivity in some sachet drinking water samples produced in Nigeria. *Iran. J. Radiation Res.* **7**(3), 151–158 (2009).
30. Benedik, L. & Jeran, Z. Radiological of natural and mineral drinking water in Slovenia. *Radiat. Prot. Dosimetry* **151**(2), 306–313. <https://doi.org/10.1093/rpd/cs009> (2012).
31. Khandaker, M. U. et al. Radiation dose to the Malaysian population via the consumption of bottled mineral water. *Radiat. Phys. Chem.* **140**, 173–179 (2017).
32. Khandaker, M. U. et al. Assessment of radiation and heavy metals risk due to the dietary intake of marine fishes (Rastrelliger kanagurta) from the Straits of Malacca. *PLoS One* **10**, e0128790 (2015).
33. Wallner, G. & Jabbar, T. Natural radionuclides in Austrian bottled mineral waters. *J. Radioanal. Nucl. Chem.* **286**, 329–334 (2010).
34. Altikulaç, A., Kurnaz, A., Turhan, S. & Kutucu, M. Natural radionuclides in bottled mineral waters consumed in Turkey and their contribution to radiation dose. *ACS Omega* **7**, 34428–34435. <https://doi.org/10.1021/acsomega.2c04087> (2022).
35. Ramadhan, R. M., Al-Salihi, A. & Khalaf, A. H. Measurement of radiological baseline data of the hazard radiation in bottled drinking water samples in Basrah/Iraq. *Basrah J. Sci.* **38**(3), 544–558 (2020).
36. UNSCEAR. United Nations Scientific Committee on the effects of Atomic Radiation. In *Sources and Effects of Ionizing Radiation Report to the General Assembly with Scientific Annexes* (United Nations Publication, 2008).

## Author contributions

Fatemeh Ranjbar: Methodology, Investigation, Data curation, Writing–review & editing, Conceptualization, Validation, and Methodology. Hossein Sadeghi: Methodology, Investigation, Data curation, writing, review and editing, Supervision. Reza Pourimani: Investigation, Data curation, Conceptualization, Supervision. Soraya Khanmohammadi: Writing, review, and editing, Conceptualization, Validation, Software, Methodology.

## Funding

No funding was received to assist with the preparation of this manuscript.

## Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to H.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025