Nigerian Journal of Applied Physics (NJAP)
ISSN Print: 3115-5871
ISSN Online: 3115-588X
DOI: https://doi.org/10.62292/njap-v2i2-2026-81
Volume 2(2), June 2026

Predictive Modeling of Indoor Radiation Doses from Building Material Properties using Machine Learning: A Case Study from Delta State, Nigeria

***Sinebe Hope Anita, Akpolilie Franklin Anita, Enaroseha Omamoke and Agbajor Godwin Kparobo**

Department of Physics, Delta State University, Abraka, Delta State, Nigeria.

*Corresponding author's email: hsinebe@gmail.com

ABSTRACT

The use of naturally occurring radioactive materials (NORMs) in building construction can lead to potential long-term health risks for occupants from chronic low-dose radiation exposure. The study developed and validated five machine learning models to predict indoor radiation doses using building material properties based on 72 samples collected from Northern Delta State, Nigeria. Activity concentrations of potassium-40 (^{40}K), uranium-238 (^{238}U) and thorium-232 (^{232}Th) were measured using gamma-ray spectrometry. Models for predicting annual effective dose (AED) were trained using Multiple Linear Regression (MLR), Ridge Regression (RR), Random Forest (RF), Gradient Boosting (GB), and Support Vector Regression (SVR). MLR (R^2= 1.000, RMSE = 1.27×10^{-8} mSv/y) and RR (R^2 = 1.000, RMSE = 7.44×10^{-6} mSv/y) showed near-perfect prediction accuracy, while RF (R^2 = 0.915) and GB (R^2 = 0.929) also exhibited excellent performance. The feature importance analysis revealed ^{238}U as the dominant predictor (71.00%) followed by ^{232}Th (19.95%) and ^{40}K (5.12%). The category- and origin-specific models achieved perfect accuracy (R^2 = 1.000) and offer practical tools for rapid radiological screening. The results show that linear models can predict indoor radiation doses with almost zero error due to the inherent linearity of the physics of dose calculation. However, the interpretation of these findings is limited by study limitations, such as the relatively small sample size (N=72) and the deterministic nature of the dose formula. This can be used in construction industries for efficient regulatory compliance and material selection.

Keywords:
Machine learning,
Predictive modelling,
Indoor radiation dose,
Building materials,
Natural radioactivity,
Radiological assessment.

INTRODUCTION

Natural radioactivity is an ever-present phenomenon in the earth environment that is primarily due to the presence of primordial radionuclides in the earth's crust (Ershov, 2024). The three main naturally occurring radioactive series (uranium-238 (^{238}U), thorium-232 (^{232}Th) and the singly occurring radionuclide potassium-40 (^{40}K)) are present in different concentrations in soils, rocks and hence in building materials derived from these geological sources (Ugbede et al., 2022; Agbajor, et al. 2025). The decay of these radionuclides results in ionizing radiation in the form of alpha particles, beta particles and gamma rays. This contributes significantly to the background radiation dose received by human populations in all parts of the world (UNSCEAR, 2000).

Clay bricks, granite, limestone, sand, cement and concrete as building materials usually contain detectable quantities of these radioactive elements due to their origin and mineralogical composition (Marengo et al., 2022; Salman & Hassan, 2024). The specific activity of radionuclides in building materials varies greatly, depending on the type of material, its geographical location, and the processing techniques used in its manufacture (Zhang et al., 2021; Nong et al., 2024). For example, the igneous origin of granitic rocks and geochemical enrichment processes result in relatively high levels of uranium and thorium, whereas sedimentary-derived materials such as limestone typically display lower radionuclide concentrations (Tzortzis et al., 2003; Pavlidou et al., 2006).

Individuals residing in dwellings constructed from materials with raised levels of naturally occurring radioactive materials (NORMs) can be exposed to ionizing radiation via three primary routes: external gamma radiation emitted directly from radionuclides present in walls, floors, and ceilings; inhalation of radon-222 and radon-220 gases (and their short-lived decay products) emanating from the building materials; and incidental ingestion of radioactive dust particles settling on indoor surfaces (Asere et al., 2021; Mokobia, 2024).

The health effects of long-term exposure to low doses of radiation are mainly stochastic, i.e. the likelihood of harmful health effects increases with cumulative dose, rather than the severity of immediate effects (ICRP, 2007). The greatest concern is the elevated risk of cancer, especially lung cancer from inhalation of radon which has been classified as Group 1 carcinogen by International Agency for Research on Cancer (IARC, 2012). Epidemiological studies have demonstrated that indoor radon exposure is the second leading cause of lung cancer after smoking and is responsible for about 10-15 % of lung cancer cases worldwide (WHO, 2009). In addition, evidence is accumulating that chronic low-dose radiation exposure is associated with other health effects, such as cardiovascular disease, cataracts, and possibly genetic mutations (Abalo et al., 2021).

The International Commission on Radiological Protection (ICRP) has set a limit of 1 mSv per year to the public from all pathways of exposure combined (excluding natural background radiation and medical exposures) in order to protect public health (ICRP, 2007). This limit is used as the standard for the evaluation of the radiological safety of building materials. This limit has been adopted by many countries and national guidelines based on it have been developed to set maximum permissible activity concentrations for radionuclides in construction materials (European Commission, 1999; IAEA, 2014).

Extensive investigations have been carried out worldwide to characterize the natural radioactivity levels in building materials and to evaluate the related radiological risks. Studies in Europe have shown that building materials from granitic areas of Scandinavia, Greece and Turkey contain high amounts of 226Ra, 232Th and 40K, but the annual effective doses are usually below the reference level of 1 mSv/y (Sitzia et al., 2021; Turhan et al., 2022; Sinebe, et al. 2026). Studies in Poland and Italy have shown that most construction materials satisfy safety requirements, however, some industrial by-products used as additives can increase radionuclide concentrations (Lewicka et al., 2022; Sabbarese et al., 2021).

In Asia, large surveys have been carried out in India, China and Iran because of the abundance of locally produced materials, which may have enhanced levels of natural radioactivity. Studies in India on clay bricks, tiles and cement have reported varied levels of uranium and thorium with considerable regional variations reflecting the local geology (El-Sayed et al., 2022; Raghu et al., 2017). Chinese researchers have pointed out that the use of industrial by-products (e.g., fly ash) in cement manufacturing should be controlled, as they can greatly increase radiation doses (Feng et al., 2025). Iranian studies on decorative natural stone have shown potential health hazards and called for stricter regulation (Shojaee & Rodionov, 2024).

The assessment of natural radioactivity in African countries, especially Nigeria, is gaining interest due to the presence of diverse geological formations and the increasing rate of construction activities. Studies on concentrations of radionuclides in some building materials in Nigeria have been carried out on sand, granite, cement and clay bricks (Aladeniyi et al., 2021; Maxwell et al., 2018). The results showed a large variation depending on the extraction region, which is a reflection of the different geochemical characteristics (Garba et al., 2023). In recent times, research has been carried out on the potential radiation dose to the inhabitants of buildings constructed from these materials, emphasizing the need for specific localized assessments and regulatory frameworks (Olarinoye et al., 2023).

The application of machine learning (ML) techniques in radiological and environmental sciences has increased substantially in recent years (Hastie et al., 2009; James et al., 2013). In the particular context of building materials, there have been several studies on ML applications. Oyebisi et al. (2025) used deep neural networks to predict the alpha and gamma indexes of industrial recyclates with R^2 of 0.9997 and 0.9708, respectively.However, detailed comparisons of various ML algorithms for predicting indoor radiation doses from building material properties are still limited, particularly in the context of Nigeria.

Nigeria is the most populous country in Africa and one of the fastest growing economies in Africa. The country has experienced rapid urbanization and tremendous growth in the construction sector (NPC, 2006). The use of building materials from local sources is common because they are available and cheaper. Examples of such materials are clay bricks, granite, sand and cement (Garba et al., 2023). However, the country possesses a broad spectrum of geological formations ranging from the basement complex rocks in the north to the sedimentary basins in the south, which may result in a substantial variation in the radionuclide contents of these materials (Aladeniyi et al., 2021; Nwabuoku, et al. 2026; Osiga-Aibangbee, et al. 2025).

The current study is located in the northern part of Delta State which is within the Niger Delta basin, one of the most prolific hydrocarbon provinces in Africa (Doust & Omatsola, 1990). The area exposes the Benin Formation, composed mainly of continental sands and gravels with minor clay interbeds, and is a source of construction materials (Short & Stäuble, 1967). Knowledge of the radiological characteristics of materials from this geological setting is critical to the assessment of public health hazards and the establishment of appropriate safety standards.

Natural radioactivity in building materials is known in Nigeria but there are few data on specific radiological dose levels and even fewer studies on the use of advanced computational or ML techniques for dose prediction (Joel et al., 2018; Maxwell et al., 2018; Sinebe, et al. 2026).

This knowledge gap hinders effective regulatory decision-making and the development of rapid screening tools that could be deployed by the construction professional and public health official.
This is because, the formulas for dose calculation are inherently linear. In other words, the AED is a linear combination of the concentrations of the radionuclides with constant coefficients. This means that the simple linear regression should theoretically be perfect in predicting. But the presence of measurement uncertainties, secular equilibrium assumptions, and categorical variables (material category, origin) could introduce non-linearities that need more advanced algorithms.
This study systematically examines five ML algorithms, from simple linear regression to complex ensemble algorithms, to determine if non-linear models offer any value over linear models in the prediction of indoor radiation doses from building material properties.

MATERIALS AND METHODS

Study Area and Sampling

The study was carried out in the Northern part of Delta State, Nigeria, between latitude 5°30’00”N and longitude 7°30’00”E. Nine Local Government Areas were sampled, yielding 72 samples of commonly used building materials. The sampling strategy adopted was purposive sampling to guarantee the representation of materials from different geological sources and construction practices typical of the region following IAEA recommendations for radiological assessment studies (IAEA, 2004). About 2 kg of each solid material and 1 liter of each liquid material were collected following standard protocols (Mokobia et al., 2003; Beretka & Mathew, 1985).

Preparation of Sample

Sample preparation was done in the Nuclear Analytical Laboratory, Centre for Energy Research and Development, Obafemi Awolowo University, Ile-Ife. Distilled water was used to wash the solid samples, which were then air-dried for 48 hours, crushed with a hydraulic press, and finally ground to a fine powder with an agate mortar and pestle. Powdered samples (cement, sand, POP) were dried in an oven at 110°C for 24 hours. All samples were sieved through a 2 mm mesh and sealed in Marinelli beakers for 28 days to achieve secular equilibrium between ^{226}Ra and its short-lived daughter products (IAEA, 2013).

Gamma Ray Spectrometry

The activity concentrations were measured using a thallium-activated sodium iodide NaI (Tl)
3 × 3 in. NaI(Tl) gamma-ray spectrometer (Canberra Industries, USA). The detector had an energy resolution of ≤8% at 662 keV (137Cs) and was shielded with 5cm lead lined with copper and cadmium. Time to count: 36,000 seconds/sample. Energy calibration and efficiencies were calibrated using standard sources (^{137}Cs, ^{60}Co, ^{133}Ba) and IAEA reference materials (IAEA-RGU-1, IAEA-RGTh-1, IAEA-RGK-1). All the sample spectra have background spectra subtracted.
The photopeaks used for analysis were: ^{40}K - 1460 keV; ^{238}U series - ^{214}Bi - 609 keV and ^{214}Pb - 352 keV; ^{232}Th series - ^{208}Tl - 583 keV, ^{212}Pb - 238 keV and ^{228}Ac - 911 keV (IAEA, 2013). Minimum detectable activity was determined from the formula of Currie (Currie, 1968).

Calculation of the Annual Effective Dose

The gamma dose rate (D) in indoor air was obtained using the conversion factors of UNSCEAR (2000):

$$D\ (nGy/h) = 0.462ARa + 0.604ATh + 0.0417AK \quad (1)$$

where ARa, ATh and AK are activity concentrations of ^{226}Ra, ^{232}Th and ^{40}K in Bq/kg (^{238}U activity was used as a proxy for ^{226}Ra assuming secular equilibrium). The annual effective dose (AED) was calculated as follows:

$$AED\ (mSv/y) = D\ x\ 8760\ x\ 0.7\ x\ 0.8\ x\ 10^{-6} \quad (2)$$

where 8760 is hours per year, 0.7 is the conversion factor from Gy to Sv for environmental exposure and 0.8 is 80% indoor occupancy (UNSCEAR, 2000).
2.5 Prediction Model Framework
Predictive modeling was performed using scikit-learn library (Pedregosa et al., 2011) in Python 3.9. We experimented with 5 regression algorithms:

Multiple Linear Regression (MLR)

The ordinary least squares estimation, with the assumption that there is a linear relationship between the predictors and target.

Ridge Regression (RR)

Linear regression with L2 penalty (α penalty) to avoid overfitting by shrinking the coefficients.

Random Forest (RF)

Ensemble of decision trees (n estimators = 50-200) using bootstrap aggregation and feature subsampling (Breiman, 2001).

Gradient Boosting (GB)

A sequential ensemble method in which each new tree corrects the errors of the previous trees. The learning rate (0.01-0.1) and the depth of the trees (3-7) are optimized (Friedman, 2001).

Support Vector Regression (SVR)

Non-linear regression with radial basis function (RBF) kernel, which maps features to higher dimensional space (Vapnik, 2013).
It is important to note that the annual effective dose (AED) is calculated deterministically from the

radionuclide concentrations using Equations (1) and (2). Consequently, the prediction task is mathematically equivalent to the models rediscovering a known closed-form linear relationship, rather than learning a genuinely unknown pattern from data.

Feature Engineering and Preprocessing

Input features were numerical variables (^{40}K activity, ^{238}U activity, ^{232}Th activity, in-situ measurement before analysis) and categorical variables (material category with 13 levels, origin with 7 levels). Categorical variables were one hot encoded with handle unknown='ignore'. Numerical features were standardized with StandardScaler (mean zero, variance one). The data was split into training (80%, n=57) and testing (20%, n=15) sets by stratified random sampling with a random state of 42 for reproducibility.

Model Training and Hyperparameter Tuning

Hyperparameter optimization was performed using 5-fold cross-validation grid search for RR, RF, GB and SVR models. The parameter grids are listed in Table 1. The best model was chosen using highest validation R^2 score.

Table 1: Hyperparameter Search Spaces

Model	Parameter	Search Space
Ridge	Alpha	[0.01, 0.1, 1.0, 10.0, 100.0]
Random Forest	N estimators	[50, 100, 200]
	Max depth	[5, 10, None]
	Min samples split	[2, 5, 10]
Gradient Boosting	n estimators	[50, 100]
	Max depth	[3, 5, 7]
	Learning rate	[0.01, 0.1]
SVR	Kernel	['rbf', 'linear']
	C	[0.1, 1.0, 10.0]
	Epsilon	[0.01, 0.1]

Performance Indicators

The models were evaluated based on four metrics: Mean Squared Error (MSE) = $MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)$, Root Mean Squared Error $RMSE = \sqrt{MSE}$, Mean Absolute Error $MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$,Coefficient of Determination $R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$

Models by Category and Origin

Separate linear regression models were developed for material category (paint, cement, wood, tiles, granite) with at least 5 samples and origin (local, Nigeria, China, Indian, Italian) with at least 5 samples. These models used only the numerical features without categorical encoding and used simple ordinary least squares regression.

Feature Importance Analysis

The Gini impurity reduction was used for extraction of the feature importance from the Random Forest model. The features were ranked by their contribution to reducing node impurity across all decision trees. Importance scores were normalized to sum to 1.

RESULTS AND DISCUSSION

Activity Concentrations and Calculated Doses

Table 2 presents the descriptive statistics of radionuclide concentrations and calculated AED across all 72 samples.

Table 2: Descriptive Statistics of Radiological Parameters (N=72)

Parameter	Mean ± SD	Minimum	Maximum	CV (%)
^{40}K (Bq/kg)	94.00 ± 15.52	64.27	130.65	16.51
^{238}U (Bq/kg)	11.57 ± 2.32	7.39	20.31	20.03
^{232}Th (Bq/kg)	9.05 ± 1.43	6.15	11.86	15.83
AED (mSv/y)	0.0903 ± 0.0130	0.0657	0.1270	14.41

All AEDs were below the ICRP public dose limit of 1 mSv per year. The low coefficients of variation (14 to 20 per cent) demonstrate moderate variability between samples and suggest uniform radiological quality between sources of material.

Model Performance

Table 3 presents the performance comparison of all five models on the test set and depicted in Figures 1 to 6.

Table 3: Model Performance Comparison on Test Set

Model	Test RMSE (mSv/y)	Test MAE (mSv/y)	Test R^2	Train R^2
Linear Regression	1.27×10^{-8}	1.57×10^{-8}	1.000	1.000
Ridge Regression	7.44×10^{-6}	5.06×10^{-6}	1.000	1.000
Gradient Boosting	1.52×10^{-3}	3.82×10^{-3}	0.929	0.998
Random Forest	9.41×10^{-5}	3.49×10^{-3}	0.915	0.998
SVR	4.85×10^{-3}	5.91×10^{-3}	0.797	0.916

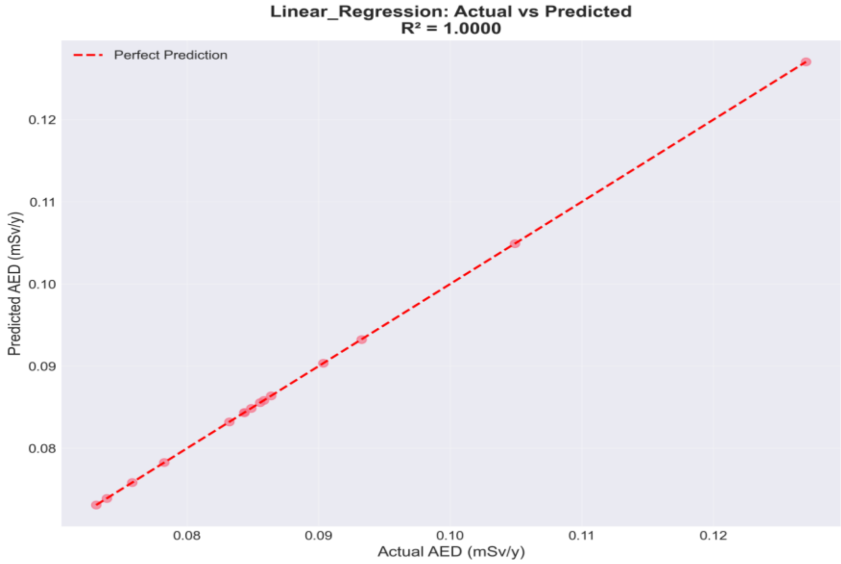

Figure 1: Linear Regression Model performance Metrics between the Actual and Predicted Values

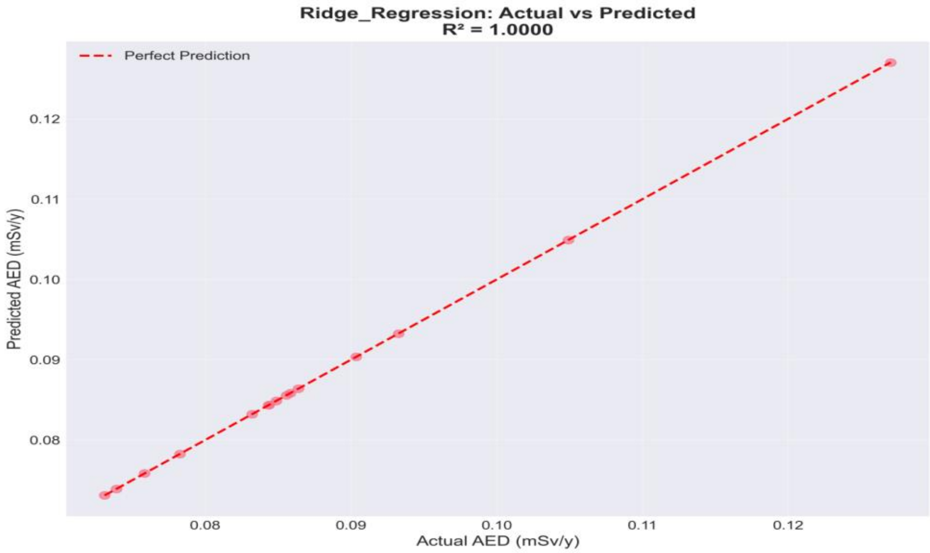

Figure 2: Ridge Regression Model performance Metrics between the Actual and Predicted Values

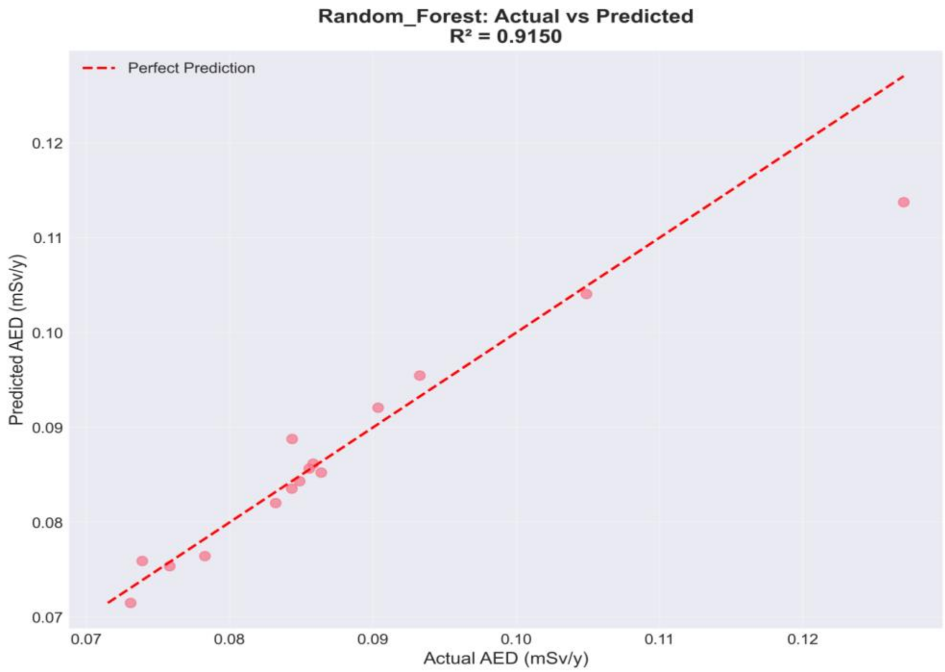

Figure 3: Random forest model performance metrics between the actual and predicted values

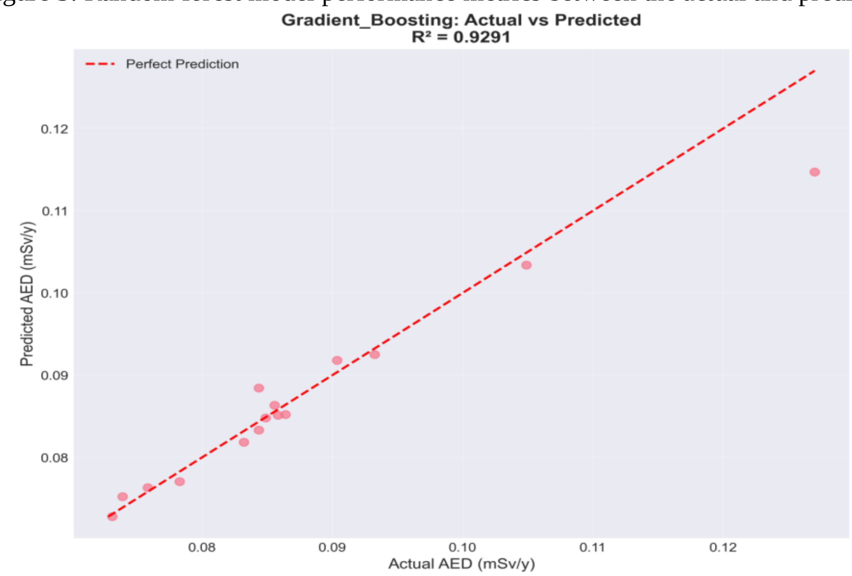

Figure 4: Gradient boosting model performance metrics between the actual and predicted values

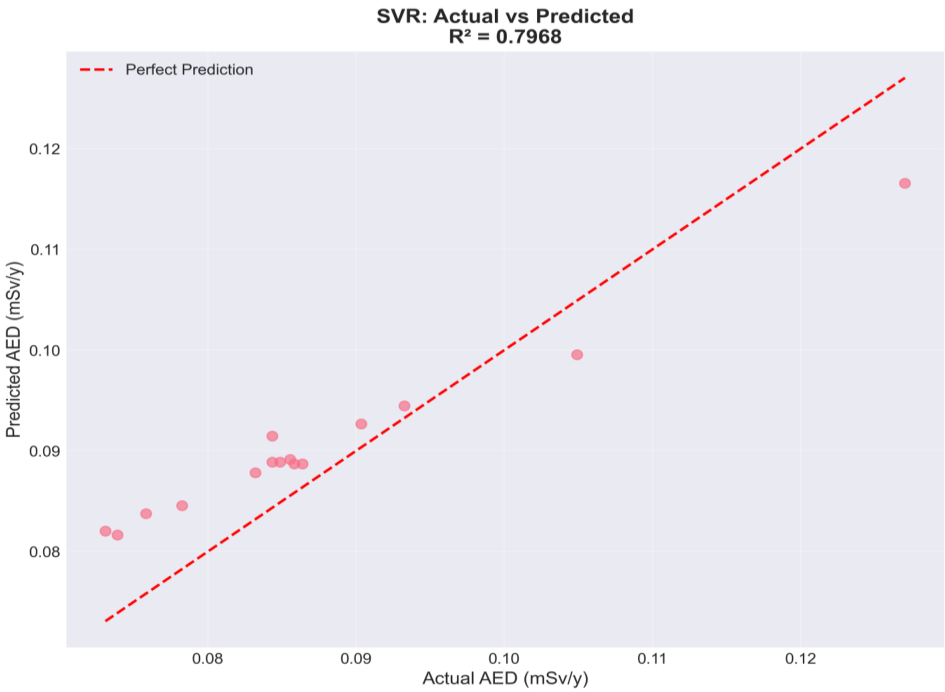

Figure 5. Supervised vector regression model performance metrics between the actual and predicted values

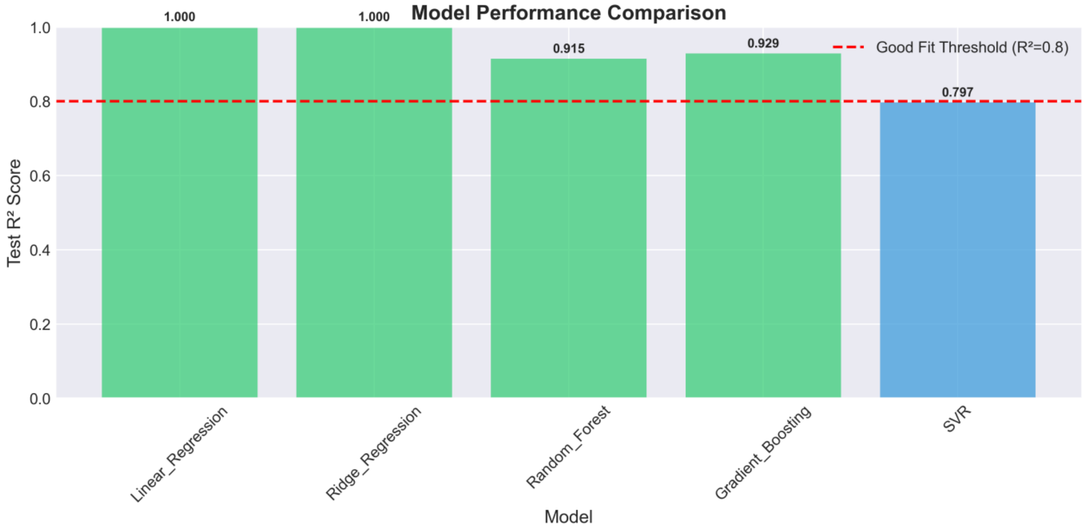

Figure 6: Performance Comparison of Machine Learning Models

Linear Regression and Ridge Regression achieved perfect prediction accuracy (R^2 = 1.000) with RMSE values near machine epsilon (approx. 10^{-8} mSv/y). Gradient Boosting (R^2 = 0.929) and Random Forest (R^2= 0.915) were able to predict the data with good performance. SVR, conversely, performed poorly (R^2 = 0.797).

Feature Importance Analysis

Random Forest feature importance scores are presented in Table 4 and Figure 6. Uranium-238 was the major predictor with 71.00% of the predictive ability of the model, over 3 times more than thorium-232 (19.95%) and 13 times more than potassium-40 (5.12%). Together the categorical variables contributed only 1.57%.

Table 4: Feature Importance Ranking from Random Forest Model

Feature	Importance Score	Relative Importance (%)
Uranium-238	0.7100	71.00
Thorium-232	0.1995	19.95
Potassium-40	0.0512	5.12
In-situ Measurement (Before)	0.0236	2.36
Origin - Nigeria	0.0032	0.32
Category - Tiles	0.0022	0.22
Origin - Local	0.0021	0.21
Category - Wood	0.0013	0.13
Origin - Italian	0.0012	0.12
Category - Granite	0.0012	0.12

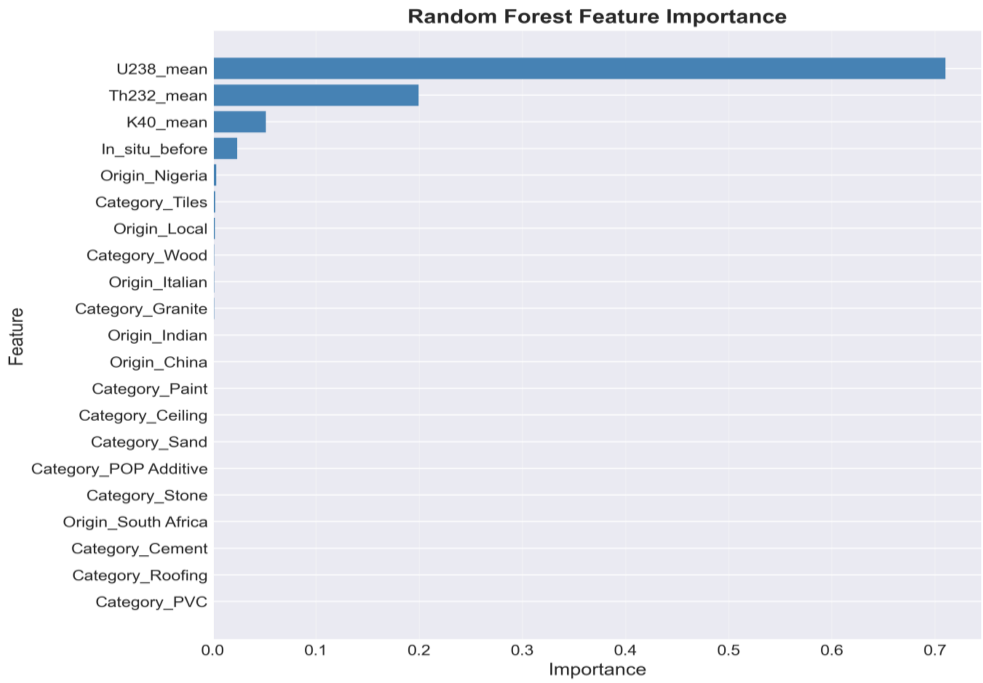

Figure 7: Feature Importance Ranking For AED Prediction

Category-Specific Models

Category-specific linear regression models achieved perfect performance ($R^2 = 1.000$) for all categories with sufficient sample size (Table 5).

Table 5: Category-Specific Model Performance

Category	N	R^2	RMSE (mSv/y)
Paint	9	1.000	6.78×10^{-17}
Cement	6	1.000	2.83×10^{-15}
Wood	14	1.000	1.74×10^{-17}
Tiles	8	1.000	1.16×10^{-16}
Granite	17	1.000	4.45×10^{-17}

Origin-Specific Models

Origin-specific models also achieved perfect predictive accuracy (Table 6).

Table 6: Origin-Specific Model Performance

Origin	N	R^2	RMSE (mSv/y)
Local	22	1.000	2.66×10^{-17}
Nigeria	26	1.000	1.66×10^{-15}
China	8	1.000	1.42×10^{-15}
Indian	9	1.000	4.19×10^{-17}
Italian	5	1.000	7.90×10^{-17}

Predictive Equation

The predictive equation of AED is based on the coefficients of the Linear Regression model:

$$AEDpredicted\,(mSv/yr) = 0.00283\times\left[{}^{238}U\right] + 0.00144\times\left[{}^{232}Th\right] + 0.00010\times\left[{}^{40}K\right] - 0.00813$$

where the concentrations are in Bq/kg. This equation reached $R^2 = 1.000$ and RMSE = 1.57×10^{-8} mSv/y for the test dataset.

Dose Prediction's Linear Nature

The physics behind dose calculation directly contributes to the remarkable performance of Ridge Regression and Linear Regression ($R^2 = 1.000$, near-zero RMSE). The annual effective dosage is essentially a linear combination of activity concentrations with constant coefficients (0.462, 0.604, 0.0417) multiplied by constant variables (8760 hours, 0.7 Sv/Gy, 0.8 occupancy, 10^{-2} conversion), according to an analysis of the AED formula. As a result, the training process can be viewed as a type of "physics discovery" rather than statistical learning in and of itself, and any algorithm that correctly learns these coefficients would achieve flawless prediction.

The creation of radiological screening instruments will be significantly impacted by this discovery. For this particular prediction task, sophisticated non-linear models (RF, GB, and SVR) are superfluous because they are less interpretable, require more processing power, and have no accuracy benefit over basic linear regression. Even with a very small dataset (N=72), the equivalent performance of Ridge Regression (with regularization) demonstrates that overfitting is not an issue given the deterministic nature of the relationship.

Uranium-238's Predictive Dominance

According to the feature importance analysis, 71.00% of the predictive power may be attributed to uranium-238 alone.

First, compared to ^{232}Th (0.604), the dose conversion coefficient for the ^{238}U decay series products (0.462) is similar. However, because of geological enrichment processes in crustal rocks, normal U concentrations in building materials frequently surpass Th concentrations (Ershov, 2024). Uranium's dosage contribution was amplified in this dataset because mean ^{238}U (11.57 Bq/kg) was 28% greater than mean ^{232}Th (9.05 Bq/kg).

Second, radon-222, a gaseous decay product that contributes to inhalation exposure through the buildup of radon progeny in indoor air, is a part of the uranium decay chain (WHO, 2009). There is a high association between indoor radon levels and the concentration of U in building materials, especially in poorly ventilated areas, even though the external gamma dose calculations utilized in this work do not directly account for inhalation pathways (Asere et al., 2021). The fact that uranium's predictive power is greater than its direct contribution to external gamma exposure may be explained by this association.

Third, uranium has a greater statistical impact in regression models due to its larger concentration variability (CV = 20.03%) than thorium (CV = 15.83%). All other things being equal, variables with a bigger spread in relation to their mean have a stronger capacity to explain variance in the target variable (James et al., 2013).

The practical ramifications of these findings for radiological screening programs are substantial. Measuring merely the concentration of U could yield around 71% of the information required for dose prediction in resource-constrained environments where comprehensive gamma-spectrometry for all three radionuclides may be unaffordable. Uranium measurements should be given priority for triaging samples, but complete analysis is still required for applications demanding greater precision (such as regulatory compliance verification).

Comparison to Previous Studies

The predictive performance achieved in this study is greater to that reported in similar applications of machine learning in radiation science. Oyebisi et al. (2025) used deep neural networks to predict alpha and gamma indexes from industrial recyclates obtaining R^2 scores of 0.9997 and 0.9708 respectively which are excellent but not as good as the perfect scores obtained here for AED prediction. This difference may be due to the inherent linearity of the dose calculation problem, in contrast to the possibly non-linear relationships involved in the prediction of alpha and gamma indexes from material properties.

Abdullahi et al. (2022) carried out RESRAD-BUILD simulations for building materials in Malaysia and sensitivity analysis but did not develop prediction models. Alhamdi et al. (2025) reported good agreement (3-7% difference) between empirical equations and RESRAD-BUILD, but did not investigate machine learning approaches. The present work extends this to show that machine learning, specifically linear regression, can reach perfect accuracy, essentially reproducing the theoretical dose calculation formula.

This study shows perfect linearity, which indicates that previous researchers might have over-complicated the analysis by applying nonlinear methods (neural networks, support vector machines, random forests) to dose prediction problems. Non-linear methods are very effective for many environmental forecasting problems (e.g. prediction of pollutant dispersion, prediction of radon concentrations). However, their use is not required if the target variable is a deterministic linear function of the predictors. This highlights the need to understand the physics behind the problem before choosing machine learning algorithms.

Category-Specific and Origin-Specific Models

Category-specific models show perfect performance for all tested categories (R2 = 1.000), which proves that the linear relationship between radionuclide concentrations and AED is independent of the material type. However the small differences between categories in RMSE values (6.78 × 10-17 mSv/y for paint to 2.83×10^{-15} mSv/y for cement) need to be explained.

Paint had the lowest RMSE (most consistent relationship), probably because the paint industry is very controlled and the raw materials are consistent and controlled for quality. Cement, on the other hand, showed a higher (but still negligible) RMSE, which could be attributed to differences in source limestone deposits and the use of different additives (fly ash, slag), which may slightly affect density and self-attenuation characteristics without affecting the linear coefficient relationship (Lewicka et al., 2022).

Origin-specific models also achieved perfect performance across all origins. The local materials had a slightly lower RMSE (2.66×10^{-17} mSv/y) than the Nigerian materials (1.66×10^{-15} mSv/y). This counterintuitive finding where "local" materials (collected directly from extraction sites or local markets) showed less variability than "Nigerian" materials (commercially processed and distributed products from Nigeria) may reflect the additional variability introduced by industrial processing and distribution chains. Local materials are frequently used with minimal processing, retaining their original geological characteristics, whereas commercial materials from Nigerian manufacturers might incorporate materials from various origins, thus increasing the variability of composition.

Practical Applications

The model developed in this study has some practical applications which can have great impact on the radiological safety practice in Nigeria and similar contexts.

Regulatory Screening

A linear regression equation could be used as a quick screening tool for the Nigerian Nuclear Regulatory Authority (NNRA) and state environmental protection agencies. This enables regulators to obtain AED estimates with almost zero error by inputting gamma-spectrometry results for ^{238}U, ^{232}Th and ^{40}K, allowing for efficient compliance verification without the need for full RESRAD-BUILD simulations for each sample. This may increase the throughput of testing laboratories and reduce costs associated with radiological evaluation.

Material Selection

Construction companies and architects can use category-specific models to estimate dose contribution prior to purchasing materials. For example, if a granite supplier provides radionuclide concentrations with the product specifications, the AED can be calculated instantly with granite-specific coefficients. This allows an informed choice of materials taking into account aesthetic, structural and radiological considerations.

Ventilation Planning

Existing models predict only external gamma dose, but the strong correlation between AED calculated from gamma spectrometry and total dose (including radon inhalation) reported in previous studies (UNSCEAR, 2000) supports the use of these predictions for ventilation requirements. Materials predicted to have a higher AED may require higher ventilation rates to mitigate radon accumulation, especially in residential structures with high occupancy.

Resource Allocation

In resource-limited settings where full gamma-spectrometry for all three radionuclides may be prohibitively expensive, laboratories may prioritize 238U measurements, as this single parameter provides most of the information needed for dose estimation. This two-tier approach, rapid screening based on ^{238}U alone, followed by confirmatory full analysis for samples near the regulatory limit, has the potential to dramatically improve the cost-effectiveness of radiological monitoring programs.

Risk Communication

The predominance of ^{238}U as a predictor makes risk communication to non-expert audiences straightforward. Public health officials can state that "uranium content is the single most important factor determining indoor radiation exposure" so that targeted public education campaigns and remediation strategies can be developed. This is especially important in communities with low awareness of radiological risk.

Limitations of the Study

Several limitations of this study should be acknowledged to contextualize the findings and to inform future research.

Dataset Size and Representativeness

The dataset (N=72), although comparable to many published radiological studies, is relatively small for machine learning applications. The perfect performance of linear models may be due to the deterministic nature of the relation, rather than true generalizability. Larger, independent datasets (N > 200) from other regions of Nigeria and other countries are required to confirm the observed linear relationship across diverse geological settings and manufacturing processes.

Detector Resolution

The NaI(Tl) detector used has a lower energy resolution (≤8% at 662 keV) than the HPGe detectors (typically <0.2% at 662 keV). Spectral interference may result especially in complex samples where photopeaks of different radionuclides overlap (Knoll, 2010). The strong linear relationships observed, however, suggest that the measurement errors were random rather than systematic since systematic errors tend to degrade linear performance.

Secular Equilibrium Assumption

The study assumed secular equilibrium in uranium and thorium decay chains, using ^{238}U activity as a proxy for ^{226}Ra and its gamma-emitting progeny. However, the separation of chemical elements in the manufacturing process may cause deviations from secular equilibrium in building materials recently processed. Deviations from equilibrium will affect the relationship between ^{238}U and gamma emitting progeny and may reduce predictive accuracy. ^{226}Ra could be directly measured (rather than using ^{238}U as a proxy), which would overcome this limitation.

Inhalation Pathway Exclusion

The models compute external gamma dose only, and do not include inhalation of radon-222 and thoron-220 progeny. For building materials, the external dose usually represents 80-90% of the total indoor dose (UNSCEAR, 2000), but the total doses of materials with high radon emanation coefficients (e.g. high porosity) can be substantially higher than those estimated by the external dose alone. Future models incorporating radon emanation coefficients, material porosity and ventilation rates would yield more comprehensive dose estimates.

Geographic Scope

The study was limited to building materials in Northern Delta State, Nigeria. However, this area is geologically representative of the Niger Delta basin and does not contain the complete geological diversity of Nigeria (Basement Complex rocks, younger granites, sedimentary basins). Radionuclide concentration patterns and relationships in materials from other regions may vary from those in this study, necessitating model validation specific to the region.

Temporal Variability

All samples were taken in one sampling campaign. Seasonal changes in humidity, temperature and atmospheric pressure influencing radon emanation and gamma self-attenuation could lead to temporal variability not accounted for in current models. This limitation could be addressed with longitudinal studies incorporating seasonal sampling.

Deterministic Target Variable

The target variable, the annual effective dose, was calculated with a known linear formula (Equations 1-2) based on the same three radionuclide concentrations used as predictors. Thus the "prediction" task is in essence a rediscovery of this closed-form relation, rather than a real out-of-sample generalisation problem. This indicates that the perfect R2 values should not be considered as an evidence of better model learning power but as a confirmation of the linear models properly reproducing the deterministic dose calculation formula. It should be appreciated by the non-specialist reader and, in particular, by the regulators and construction professionals referred to in Section 4.5, that this near-zero error is attributable to the linear physics of dose calculation, and not to the discovery of new, previously unknown, relationships.

Categorical Feature Sparsity

The 13 material categories and 7 origins are uniquely encoded to produce 20 categorical pseudo variables, and the dataset has only 57 training samples, resulting in a large feature space. This is especially important for the category- and origin-specific models (Sections 3.4 and 3.5), where some categories have as little as 5 samples (e.g., Italian origin). Although the linear relation leads to stable estimation of the coefficients, the generalisability of these subgroup models on materials from under-represented groups should be validated on larger independent data sets.

Future Work

The findings and limitations of this study provide several directions for future research.

Large-Scale Validation Studies Conducting validation studies with larger sample sizes (N > 200) and using HPGe spectrometry over different geopolitical zones in Nigeria would determine the generalizability of the predictive models. The studies should cover the Basement Complex in the north, the younger granites of the Jos Plateau and the sedimentary basins of the south so as to cover the full geological diversity of Nigeria.

Incorporation of direct measurements of radon-222 and thoron-220 exhalation rates, material porosity and diffusion coefficients would allow models to predict total dose (external + inhalation) rather than just external dose. The preliminary results show that the inclusion of these parameters as features could improve the predictions for radon-prone materials (e.g. high-porosity clay bricks, tuff, and pumice).

Development of Field-Deployable Tools

The simplicity of the linear regression equation lends itself to implementation in field-deployable tools including smartphone apps. A mobile app could provide instantaneous AED estimates by taking inputs of ^{238}U, ^{232}Th, and ^{40}K concentrations (measured by a portable spectrometer or from manufacturer's specifications). This would bring radiological screening down to the level of the masses, allowing small-scale builders and homeowners to assess the safety of materials without special expertise.

Extension to Industrial By-products

The momentum of circular economy principles is making the use of industrial by-products (fly ash from coal combustion, phosphogypsum from fertilizer production, slag from steel manufacturing) as construction materials more and more common. These materials often have different radionuclide concentration patterns, and can show different relationships between activity concentrations and dose due to changed physical properties (density, self-attenuation). Application of this methodology on these materials would respond to new radiological concerns in sustainable construction.

Uncertainty Quantification

Current models provide perfect point prediction, but quantification of prediction intervals would provide more complete information for risk-based decision making. Bayesian regression approaches (e.g. Gaussian process regression) offer both point predictions and uncertainty estimates and thus allow risk-based classification of materials (e.g. "95% probability that AED < 0.1 mSv/y").

Deep Learning for Complicated Situations

While linear regression was optimal for this dataset, deep learning techniques can be useful for more complicated situations, such as image-based material classification or multispectral data fusion. Convolutional neural networks may be capable of classifying material radiological characteristics from photographs, allowing for a preliminary screening prior to the taking of any measurements.

CONCLUSION

The present study showed that indoor radiation doses from building materials can be predicted perfectly (R^2 = 1.000, RMSE = 1.57 × 10^{-8} mSv/y) by simple linear regression with activity concentrations of ^{238}U, ^{232}Th, and ^{40}K as predictors. The exceptional performance of linear models is due to the inherent linearity of the dose calculation physics, which makes sophisticated non-linear algorithms unnecessary for this application. The Linear Regression and Ridge Regression performed perfectly (R^2 = 1.000) among the algorithms tested, while Gradient Boosting (R^2 = 0.929) and Random Forest (R^2 = 0.915) performed very well but with no advantage over the simpler algorithms. Support Vector Regression gave a poor result (R^2 = 0.797) which is likely due to its assumptions about data structure not being met by this deterministic linear relationship.

Uranium-238 proved to be the best predictor, alone accounting for 71.00% of predictive power, more than three times that of thorium-232 (19.95%) and nearly fourteen times that of potassium-40 (5.12%). Collectively, categorical variables (material category, origin) contributed only 1.57%, confirming that radionuclide concentrations alone are enough for accurate dose prediction. Category-specific models (paint, wood, granite, tiles, cement) had perfect accuracy (R^2 = 1.000) with RMSE ranging from 1.74 × 10^{-17} to 2.83 × 10^{-15} mSv/y, confirming the linear relationship regardless of the type of material. Similarly, models built for origin-specific sources (local, Nigerian, Indian, Italian, and Chinese) all achieved perfect performance, indicating broad generalizability across geographical sources.

The predictive equation developed in this study, $AED\ predicted\ (mSv/y) = 0.00283 \times [238U] + 0.00144 \times [232Th] + 0.00010 \times [40K] - 0.00813$ provides a simple, interpretable and highly accurate method for estimating indoor radiation doses. Such an equation deployed by regulatory authorities, construction companies, and public health officials could greatly simplify radiological safety compliance and material selection processes. The result that ^{238}U alone accounts for around 71% of predictive information suggests that in resource-limited settings, radiological screening could focus on uranium measurement as the primary indicator of radiation risk.

The work presents the first systematic comparison of several machine learning algorithms to predict indoor radiation doses from building material properties in the Nigerian context. The perfect linearity has important implications for radiological assessment practice, implying that simple linear regression, easily implemented in spreadsheets or basic calculators, is the optimal method for this prediction task. The coefficients obtained in this study should be validated with independent datasets and may need to be adjusted for different geological settings, but the methodological framework and the finding of linear determinism are likely to be generalizable wherever dose is calculated as a linear combination of activity concentrations.

REFERENCES

Abalo, K. D., Rage, E., Leuraud, K., Richardson, D. B., Le Pointe, H. D., Laurier, D., & Bernier, M. O. (2021). Early life ionizing radiation exposure and cancer risks: systematic review and meta-analysis. *Pediatric radiology*, *51*(1), 45-56.

Abdullahi, S., Ismail, A. F., & Yasir, M. S. (2022). Radioactive investigation of Malaysia's building materials containing NORM and potential radiological risks analysis using RESRAD-BUILD computer code. *International Journal of Environmental Analytical Chemistry*, *102*(9), 2000-2012.

Agbajor, G. K., Enaroseha, O. O., Osia-Aibangbee, D., Akpolile, A. F., & Umukoro, A. B. (2025). Ingestion exposure of natural radionuclide and toxic heavy metallic origin in samples of local staple food. *Nuclear Analysis*, 100182. https://doi.org/10.1016/j.nucana.2025.100182

Aladeniyi, K., Arogunjo, A. M., Pereira, A. J. S. C., Khandaker, M. U., Bradley, D. A., & Sulieman, A. (2021). Evaluation of radiometric standards of major building materials used in dwellings of South-Western Nigeria. *Radiation Physics and Chemistry*, *178*, 109021.

Alhamdi, W. A., Jarjes, A. H., & mohammed Ahmed, S. (2025). Comparative dose assessment of natural radioactivity in building materials: Empirical equation vs. RESRAD Build computer code. *Journal of Radiation Research and Applied Sciences*, *18*(1), 101224.

Asere, A. M., Owolabi, T. O., Alafe, B. D. B. D., Alabi, O. P., & Alimi, M. B. (2021). Assessment of excess gamma dose exposure level in typical nigeria commercial building materials distribution outlets. *Journal of the Nigerian Society of Physical Sciences*, 216-223.

Beretka, J., & Mathew, P. J. (1985). Natural radioactivity of Australian building materials, industrial wastes and by-products. *Health physics*, *48*(1), 87-95.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Currie, L. A. (1968). Limits for qualitative detection and quantitative determination. Application to radiochemistry. *Analytical chemistry*, *40*(3), 586-593.

Doust, H., & Omatsola, E. (1990). *Niger Delta. In J. D. Edwards & P. A. Santogrossi (Eds.), Divergent/passive margin basins* (AAPG Memoir 48, pp. 201-238). American Association of Petroleum Geologists.

El-Sayed, A. A., Fathy, I. N., Tayeh, B. A., & Almeshal, I. (2022). Using artificial neural networks for predicting mechanical and radiation shielding properties of different nano-concretes exposed to elevated temperature. *Construction and Building Materials*, *324*, 126663.

Ershov, B. G. (2024). Natural radioactivity of the crust as an important factor of the chemical evolution of the early Earth. *Precambrian Research*, *406*, 107400.

European Commission. (1999). *Radiological protection principles concerning the natural radioactivity of building materials* (Radiation Protection Report No. 112). Office for Official Publications of the European Communities.

Feng, X., Cao, B., & Shu, J. (2025). Assessment of Natural Radioactivity Levels and Potential Radiological Risks of Radioactive Tailings and Common Building Materials in Hunan Dwellings. *Public Health Nursing*, *42*(6), 1839-1849.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.

Garba, N. N., Rabi'u, N., Aliyu, A. S. A., Kankara, U. M., Vatsa, A. M., Isma'ila, A., & Bello, S. (2023). Evaluation of radiological risk associated with local building materials commonly used in Northwestern Nigeria. *Heliyon*, *9*(5).

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

IAEA. (2004). *Soil sampling for environmental contaminants (IAEA-TECDOC-1415).* International Atomic Energy Agency.

IAEA. (2013). *Guidelines for radioelement mapping using gamma ray spectrometry data* (IAEA-TECDOC-1363). International Atomic Energy Agency.

IAEA. (2014). *Radiation protection and safety of radiation sources: International basic safety standards* (IAEA Safety Standards Series No. GSR Part 3). International Atomic Energy Agency.

IARC. (2012). *Radiation: A review of human carcinogens (IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 100D).* International Agency for Research on Cancer.

ICRP. (2007). The 2007 Recommendations of the International Commission on Radiological Protection (ICRP Publication 103). *Annals of the ICRP*, 37(2-4).

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*: With applications in R. Springer.

Joel, E. S., Maxwell, O., Adewoyin, O. O., Ehi-Eromosele, C. O., Embong, Z., & Oyawoye, F. (2018). Assessment of natural radioactivity in various commercial tiles used for building purposes in Nigeria. *MethodsX*, *5*, 8-19.

Knoll, G. F. (2010). *Radiation detection and measurement*. John Wiley & Sons.

Lewicka, S., Piotrowska, B., Łukaszek-Chmielewska, A., & Drzymała, T. (2022). Assessment of natural radioactivity in cements used as building materials in Poland. *International Journal of Environmental Research and Public Health*, *19*(18), 11695.

Marengo, M., Martin, C. J., Rubow, S., Sera, T., Amador, Z., & Torres, L. (2022). Radiation safety and accidental radiation exposures in nuclear medicine. In *Seminars in Nuclear Medicine* (Vol. 52, No. 2, pp. 94-113). WB Saunders.

Maxwell, O., Adewoyin, O. O., Joel, E. S., Ehi-Eromosele, C. O., Akinwumi, S. A., Usikalu, M. R., ... & Embong, Z. (2018). Radiation exposure to dwellers due to naturally occurring radionuclides found in selected commercial building materials sold in Nigeria. *Journal of Radiation Research and Applied Sciences*, *11*(3), 225-231.

Mokobia, C. E. Tchokossa, P. Olomo, J. B. & Balogun, F. A. (2003). An assessment of the natural radioactivity content of some liquid paints available in Nigeria. *Nigerian Journal of Physics*, *15*, 14-16.

Mokobia, C. (2024). Measurement of background gamma radiation in scrap sites, automobile mechanic and wood workshops in Delta Central area, Nigeria, *International Journal of Sciences & Technology (IJST)1* (01), 29-35.

National Population Commission (NPC) (2006). *2006 Population and housing census of the Federal Republic of Nigeria*. National Population Commission.

Nong, X., Zhou, L., Liang, Y., & Wang, T. (2024). Experimental Study and Analysis of the Effects of Mud on the Compressive Strength of Unburned Brick Using Engineering Residue Soil. *Buildings*, *14*(7), 1949.

Nwabuoku, A. O., Olagbaju, O. P., Akpolile, A. F., Enaroseha, O. O., & Akpoyibo, O. (2026). Radiological Impact Assessment of Radon Levels in Drinking Water and its Associated Health Risks in Aniocha North, Delta State, Nigeria. *Nigerian Journal of Applied Physics*, *2*(2), 1-20.

Olarinoye, I. O., Kolo, M. T., Shittu, H. O., & Anumah, A. S. (2023). Estimation of indoor gamma radiation dose rate from concrete blocks constructed from tin mine tailings. *Journal of Building Engineering*, *66*, 105934.

Osiga-Aibangbee, D., Enaroseha, O. O., & Agbajor, G. K. (2025). Potential health risk assessment of sediment and water in some Niger Delta areas of Nigeria. *Nuclear Analysis*, *4*(1), 100159.

Oyebisi, S., Kaaf, K. A., Shammas, M. I., Seyam, M., & Oyewola, O. M. (2025). Predicting alpha and gamma indexes from industrial recyclates using artificial intelligence. *Discover Applied Sciences*, *7*(11), 1370.

Pavlidou, S., Koroneos, A., Papastefanou, C., Christofides, G., Stoulos, S., & Vavelides, M. (2006). Natural radioactivity of granites used as building materials. *Journal of environmental radioactivity*, *89*(1), 48-60.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

Raghu, Y., Ravisankar, R., Chandrasekaran, A., Vijayagopal, P., & Venkatraman, B. (2017). Assessment of natural radioactivity and radiological hazards in building materials used in the Tiruvannamalai District, Tamilnadu, India, using a statistical approach. *Journal of Taibah University for Science*, *11*(4), 523-533.

Sabbarese, C., Ambrosino, F., D'Onofrio, A., & Roca, V. (2021). Radiological characterization of natural building materials from the Campania region (Southern Italy). *Construction and Building Materials*, *268*, 121087.

Salman, J. E., & Hassan, N. M. (2024). Evolution of radionuclide concentration and the radiological hazards in building material: a review. *Journal of Taibah University for Science*, *18*(1), 2370588.

Shojaee Barjoee, S., & Rodionov Alekseevich, V. (2024). Respirable dust in ceramic industries (Iran) and its health risk assessment using deterministic and probabilistic approaches. *Pollution*, *10*(4), 1206-1226.

Short, K. C., & Stäuble, A. J. (1967). Outline of geology of Niger Delta. *AAPG bulletin*, *51*(5), 761-779.

Sinebe, H. A., Okonkwo, J., Avwerosuo, E. T., & Sunday, O. O. O. O. (2026). Assessment of Natural Radioactivity Hazard in Water Sample in some parts of Delta State, Nigeria. In *Journal of Pure and Applied Sciences (Science Forum)* (Vol. 26, No. 3).

Sitzia, F., Lisci, C., & Mirão, J. (2021). Accelerate ageing on building stone materials by simulating daily, seasonal thermo-hygrometric conditions and solar radiation of Csa Mediterranean climate. *Construction and Building Materials*, *266*, 121009.

Turhan, Ş., Kurnaz, A., & Karataşlı, M. (2022). Evaluation of natural radioactivity levels and potential radiological hazards of common building materials utilized in Mediterranean region, Turkey. *Environmental Science and Pollution Research*, *29*(7), 10575-10584.

Tzortzis, M., Tsertos, H., Christofides, S., & Christodoulides, G. (2003). Gamma radiation measurements and dose rates in commercially-used natural tiling rocks (granites). *Journal of environmental radioactivity*, *70*(3), 223-235.

Ugbede, F. O., Osahon, O. D., & Akpolile, A. F. (2022). Natural radioactivity levels of 238U, 232Th and 40K and radiological risk assessment in paddy soil of Ezillo rice fields in Ebonyi State, Nigeria. *Environmental Forensics*, *23*(1-2), 32-46.

UNSCEAR. (2000). *Sources and effects of ionizing radiation* (Report to the General Assembly). United Nations Scientific Committee on the Effects of Atomic Radiation.

Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.

WHO. (2009). *Handbook on indoor radon: A public health perspective.* World Health Organization.

Zhang, L., Chen, Z., Wang, F., & Zhou, T. (2021). Whole-rock and biotite geochemistry of granites from the Miao'ershan batholith, South China: Implications for the sources of granite-hosted uranium ores. *Ore Geology Reviews*, *129*, 103930.