Cleaner Waste Systems 5 (2023) 100088

Contents lists available at ScienceDirect

Cleaner Waste Systems

journal homepage: www.journals.elsevier.com/cleaner-waste-systems

Application of machine learning techniques in the prediction of excess lifetime cancer risks of agricultural byproducts used as building and construction materials

Solomon Oyebisi[a,*], Hilary Owamah[b]

[a] Department of Civil Engineering, Covenant University, PMB 1023, Km 10, Idiroko Road, Ota, Nigeria
[b] Department of Civil and Environmental Engineering, Delta State University, Abraka, Nigeria

ARTICLE INFO

Keywords:
Cancer
Machine learning algorithm
Radioactivity
Recycling
Sustainable production
Waste management

ABSTRACT

Recycling improves the circular economy and resource sustainability by using agricultural waste to create new products. However, agricultural byproducts resulting from the recycling of agricultural waste materials contain naturally occurring radionuclides with potential risks to human health and the environment. Therefore, this article provides an overview of relevant literature on radiological properties of agricultural byproducts (rice husk ash, mussel shell, palm oil clinker, and palm oil fuel ash), prioritizing their specific activities (^{226}Ra, ^{232}Th and ^{40}K). Consequently, absorbed gamma dose rates (AGDR), annual effective dose rates (AEDR) and excess lifetime cancer risks (ELCR) of the agricultural byproducts studied were determined. The specific concentrations, AGDR, AEDR, and ECLR were trained, validated, and tested using various machine learning algorithms. An evaluation of the radiological properties of all agricultural byproducts examined revealed that they pose no risk of cancer. Additionally, compared to support vector machine, regression trees, ensemble trees, Gaussian process regression, and neural networks, linear regression yielded the best performance metrics, making it the most suitable technique for predicting excess lifetime cancer risks of the surveyed agricultural byproducts.

1. Introduction

The global market for resources for the built environment is expected to double in size by 2050. Conversely, rapid industrialization and urbanization have led to rapid demand for cement and accelerated the natural resources required for its production. Despite this, cement production emits about 8 % of global carbon dioxide into the atmosphere (Andrew, 2018). This results in global warming and negative environmental impacts (Belaid, 2022). Mitigating climate change and negative environmental impacts from cement production require actionable strategies for sustainable cement production. One such strategy is the recycling of agricultural wastes as substitutes for cement, integrating societal and economic needs (Aprianti et al., 2015), and environmental needs (Ghalehnovi et al., 2022).

There is worldwide concern about the naturally occurring radionuclides (NORs) in recycled wastes used as building and construction materials, based on the Uranium/Radium ($^{238}U/^{226}Ra$) series, Thorium (^{232}Th) series and Potassium (^{40}K) as underlying radiation sources (Council of the European Union [CEU], 2014; Imani et al., 2021; Kovler, 2012; Solak et al., 2014; United Nations Scientific Committee on the Effects of Atomic Radiation [UNSCEAR], 1993, 2000). This is also evident in previous studies where ^{226}Ra, ^{232}Th and ^{40}K in building and construction materials were reported as the main sources of human exposure (Faghihi et al., 2011; Mehra et al., 2010). For example, one of the main findings of the NORM databases revealed that about 42 % of the waste produced an excessive gamma dose rate below the recommended value (1 mSv y^{-1}), with building materials having a higher proportion of about 85 % (Trevisi et al., 2018). In a similar case, Mir and Rather (2014) confirmed that natural radiation sources are responsible for 70 % of the radiation dose inhaled by the human population. As a result, building occupants and potential users are exposed to external and internal radiation from NORs emitted by direct gamma radiation from radionuclides and inhalation of radon progenies (^{222}Rn) (Bavarnegin et al., 2013; Kovler, 2012; Solak et al., 2014). And continued exposure to ^{226}Ra series, ^{232}Th series, ^{40}K isotopes and radon and its progenies irradiate the cells of the pulmonary system and distribute underlying doses to the cells or tissues of the respiratory tracks, causing leukemia and anemia. Others are liver bone, pancreatic, skin,

* Corresponding author.
E-mail address: solomon.oyebisi@covenantuniversity.edu.ng (S. Oyebisi).

https://doi.org/10.1016/j.clwas.2023.100088
Received 21 December 2022; Received in revised form 29 March 2023; Accepted 1 April 2023

lung, and kidney cancers (UNSCEAR, 1993, 2000, 2008; World Health Organization [WHO], 2009). Notably, radon inhalation accounts for 3–14 % of all lung cancers, equivalent to 20,000 deaths annually in Europe (Darby et al., 2005). Overall, resources must be evaluated for potential radioactive material so prospective users can make an informed decision about their use to mitigate possible hazards (Joel et al., 2019; Maxwell et al., 2015).

Relevant studies have reported that the excess lifetime cancer risk of some building materials surpasses the world population-weighted average by UNSCEAR (2000, 2008), indicating that potential users have an inherent risk of developing cancer (Abdullahi et al., 2019). In addition to the above findings, it has been found that the risk of cancer increases with increasing exposure time to these materials. Given the health risks of some recycled waste materials, it becomes very challenging to ignore their viable existence. From the point of view of radiation protection, the assessment of the radiological risk of all materials used for construction purposes is relevant and therefore worth examining.

Recent developments in science and technology have led to a growing interest in machine learning algorithms (MLAs) for predicting and evaluating complex data with high accuracy (Okan et al., 2022; Olthof et al., 2021; Pereira and Borysov, 2019). It operates neurons like the human brain by building a system from data sets, generating patterns, and correlating factors and groups of factors (Pereira and Borysov, 2019). The machine learning algorithm consists of input, hidden, and output layers, which include input variables, hidden neurons, and target variables obtained from the hidden layer (Pereira and Borysov, 2019). Thus, the complete learning process takes place in the hidden layer where the correlation between neurons is found. Extensive research has been conducted on the application of MLAs in civil engineering for concrete mix design and strength prediction. For instances, Mohana (2020) determined the ground granulated concrete compressive strength-based concrete using machine learning models. Based on the attained prediction accuracy, random forest model demonstrated an excellent performance for predicting the compressive strength compared to support vector machine. Yeh (1998) modeled the strength of high-performance concrete using artificial neural network (ANN). The findings revealed that a strength model based on ANN is more accurate than a model based on regression analysis. Ziolkowski and Niedostatkiewicz (2019) utilized the state-of-the-art achievements in machine learning techniques for concrete mix design. The results translated the architecture of the ANN into a mathematical equation that can be used in practical applications. Liu et al. (2022) predicted alligator and longitudinal cracking of the asphalt mix design using ML techniques. The results showed that ML techniques performed better than traditional laboratory fatigue crack testing of asphalt mixes. However, there is a scare information about the evaluation and prediction of excess lifetime cancer risks associated with the use of agricultural byproducts as building and construction materials. This is the rationale for this investigation.

This paper provides a comprehensive assessment of agricultural byproducts with particular attention to the specific activities of ^{226}Ra, ^{232}Th and ^{40}K. These radionuclides were used to determine the potential hazards from the use of studied agricultural byproducts. These hazards are external and internal absorbed gamma dose rate (AGDR), external and internal annual effective dose rate (AEDR), and outdoor, indoor, and lifetime cancer risk (ELCR). In addition, the $ELCR_{total}$ (target) variables were calculated based on seven input factors (^{226}Ra, ^{232}Th, ^{40}K, $AGDR_{in}$, $AGDR_{ex}$, $AEDR_{in}$, and $AEDR_{ex}$) using the Support Learning Machine (SVM), Neural Networks (NN), Ensemble Trees (ET), regression trees (RT), Gaussian process regression (GPR) and linear regression (LR) of MLAs. The concept of K-fold validation was employed to prevent overfitting of the data. The paper compares the performance metrics of each MLA to select the best performing algorithm for predicting the excess lifetime cancer risk of agricultural byproducts. This study underpins future studies by providing the latest research on the excess lifetime cancer risks of agricultural byproducts and provides a detailed understanding of the performance of various ML techniques in predicting the ELCR of agricultural byproducts. This would save time and resources expended on assessing excess lifetime cancer risks from agricultural waste materials. In addition, it provides a framework and data pool for predicting some agricultural byproducts that may pose a cancer risk, informing potential users of the associated risks and contributing to the achievement of Sustainable Development Goal (SDG) 3 (health and well-being), SDG 11 (Sustainable Cities and Communities), and SDG 12 (Responsible Consumption and Production).

2. Data source and collection

The process includes obtaining, screening, and assessing pertinent material. A dataset was located using a number of databases, including Science Direct, Web of Science, and Google Scholar. To maximize data collection, pertinent information was also gathered from UNSCEAR, WHO, CEU, European Commission (EC), and the Canadian Nuclear Safety Commission (CNSC). Others included the Organization for Economic Co-operation and Development (OECD), International Atomic Energy Agency (IAEA), and International Commission on Radiological Protection (ICRP) (NEA-OECD). There are numerous research available as the concept of naturally occurring radioactive in building and construction materials spreads. Using terms like 'building materials,' 'construction materials,' 'machine learning approaches,' and 'supplementary cementitious materials,' 636 papers were discovered during the initial search. The search engine was further broadened to include terms like 'agricultural waste products' and 'excess lifetime cancer risks' due to the study's emphasis on applying machine learning algorithms to estimate extra lifetime cancer risks linked to agricultural byproducts. The final number of publications was 482 from the papers. Only peer-reviewed publications were taken into account to maintain the high quality of the review (Chinnu et al., 2021). The significance of the study was taken into consideration when screening, with a focus on the activity concentrations (^{226}Ra, ^{232}Th, and ^{40}K) of agricultural byproducts. The literature was further improved based on removing extraneous information, adding papers on ELCR of building and construction materials, and excluding articles on machine learning algorithms other than SVM, NN, ET, RT, GPR, and LR approaches.

By creating targeted inquiries based on the main subject of the current study, new screening methods were created. The following standards are taken into consideration by the strategies: Are the articles primarily focused on the radioactivity of agricultural byproducts? In the publications surveyed, what categories of activity concentrations are examined? Is the primary focus of the literature ELCR of agricultural byproducts? Which relationship was used globally to calculate the ELCR of building and construction materials? Does each article predict the outcomes using a specific MLA?

A sample size of 134 pertinent peer-reviewed papers was obtained after screening. By meticulously individuating, studying, examining, and reading the publications and references, these papers were researched, confirmed, and regulated. Following this careful review, about 90 highly pertinent journals were chosen for the study. After a thorough and verifiable search, only four agricultural byproducts were seen and obtained: mussel shell (MS), palm oil clinker (POC), palm oil fuel ash (POFA), and rice husk ash (RHA).

Mussel shell is an aquacultural waste, which contains high content of calcium carbonate ($CaCO_3$) (Buasri et al., 2013). The global generation of MS is about 6–8 million metric tonnes per year, but 1.5–2 million metric tonnes is recycled (Yan and Chen, 2015). Ishak et al. (2021) reviewed the effects of mussel shell ash (MSA) as concrete mixture under sodium chloride exposure. The findings revealed that the use of mussel shell ash has the potential to be a mix in concrete to improve compressive strength, density, tensile strength, and chemical resistance compared to ordinary conventional concrete. The optimum percentage of MSA as cement mixture is between 2 % and 3 % for the compressive strength, 2–4 % for the density, 2–6 % for the tensile strength, and 1 % for sodium chloride (NaCl) concentration at 2.37 %.

Palm oil clinker is a byproduct of palm oil industry that is normally dumped as waste, causing the undesirable effects to the environment sustainability. However, the utilization of POC in concrete production not only solves the problem of disposal of this solid waste but also helps conserve natural resources. For instance, Ibrahim et al. (2017) investigated the strength and abrasion resistance of palm oil clinker pervious concrete under different curing method. In this study, natural aggregates was replaced with POC aggregates, varying from 0 to 100 wt %. The results revealed a decrease in strength and abrasion resistance with increasing POC aggregates content in the mix. However, a fully water-cured method at 25 % POC aggregates substitution favorably competed with control mix compared to 50 %, 75 %, and 100 % POC aggregates replacement.

Palm oil fuel ash a product obtained from milling of palm oil clinker, which is used as a cement substitute in construction and building sector. For example, Pone et al. (2018) examined the effects of palm oil fuel ash as a cement replacement in concrete. The results showed an improved compressive strength, especially at early age; in fact POFA specimens containing 2.5 % and 5 % POFA replacement showed greater early compressive strength than the control concrete.

Rice hush ash is an agricultural byproduct, which consists of 85–90 % silica content (Aghajanian et al., 2022). The global production of rice is approximately 770 million metric tonnes per annum, of which more than 10 % is husk (Aghajanian et al., 2022). Rice husk ash is suitable as additional cementitious material, which can be obtained by controlled or natural incineration and used with or without further processing. It possesses high pozzolanic activities and very suitable as partial replacement of cement in concrete. Findings from a previous study showed that the compressive strength, flexural strength, and tensile strength of concrete specimens with 10 % cement replacement with RHA are comparable to the control specimens (Siddika et al., 2018).

3. Assessment of radiological hazard indexes

The potential carcinogenic risks of the agricultural by-products studied were assessed based on their specific activities (^{226}Ra, ^{232}Th and ^{40}K) using Microsoft Excel, 2016 and globally acceptable equations. These hazard indices are external and internal absorbed gamma dose rates (AGDR, nGy h-1) as in Eqs. (1) and (2) (CNSC (Canadian Nuclear Safety Commission), 2012; EC, 1999; UNSCEAR, 2000). Others are external and internal annual effective dose rates (AEDR, mSv y^{-1}), using Eqs. (3) and (4) (CEU, 2014; ICRP, 1994; IAEA, 2014; UNSCEAR, 2000) and excessive outdoor and indoor lifetime cancer risks (ELCR) shown in Eqs. (5)–(7) (EC, 1999):

$$AGDR_{ex} = 0.462SR_a + 0.604S_{Th} + 0.0417S_K \tag{1}$$

$$AGDR_{in} = 1.4 \times AGDR_{ex} \tag{2}$$

where 0.462, 0.604, and 0.0417 nGy h^{-1} per Bq kg^{-1} are ^{226}Ra, ^{232}Th, and ^{40}K conversion coefficients, respectively.

$$AEDR_{ex} = AGDR_{ex} \times 24h \times 365.25d \times 0.2 \times 0.7 \times 10^{-6} = 0.00123 \times AGDR_{ex} \tag{3}$$

$$AEDR_{in} = AGDR_{in} \times 24h \times 365.25d \times 1.4 \times 0.8 \times 0.7 \times 10^{-6} = 0.00687 \times AGDR_{in} \tag{4}$$

where a dose conversion, external ($_{ex}$) occupancy, and internal ($_{in}$) occupancy factors of 0.7 sy Gy^{-1}, 80 %, and 20 % were used (UNSCEAR, 2000, 2008).

$$ELCR_{ex} = AEDR_{ex} \times L_S \times R_F \tag{5}$$

$$ELCR_{in} = AEDR_{in} \times L_S \times R_F \tag{6}$$

$$ELCR_{total} = (AEDR_{ex} + AEDR_{in}) \times L_S \times R_F \tag{7}$$

Where L_S is the life span (70 years) and R_F is the risk factor (Sv^{-1}), fatal cancer risk per Sievert. For stochastic effects, ICRP 60 uses the detriment coefficient of 5.0×10^{-2} Sv^{-1} for the whole population (ICRP, 1990).

4. Machine learning algorithms

To date, scientists have used traditional computational techniques to encode a model problem based on a mathematical framework, logical inference, and established correlations within the bounded data structure (Cohen, 2021). This limited data set is tested without informing or modifying the algorithm. This rule-based technique is analogous to science, which tests hypotheses. But unlike the traditional approach, today machine learning is a hypothesis-generating technique that is evolving and emerging in science, technology, engineering, medicine, and management. It processes large and complex input data and models its internal state (hidden neurons), generating an accurate prediction (output) about the data. In short, the machine learning algorithm (MLA) is a technique that uses a standard computer architecture to model a set of input variables and derive the desired output (Cohen, 2021; Kim, 2017). These input parameters are defined as training data since the model is trained on the input variables (Thilakarathna et al., 2020). After model generation, it was tested with new and untrained data. For this study, ^{226}Ra (Bq kg-1), ^{232}Th (Bq kg-1), ^{40}K (Bq kg^{-1}), $AGDR_{ex}$ (nGy h^{-1}), $AGDR_{in}$ (nGy h^{-1}), $AEDR_{ex}$ (mSv y^{-1}), and $AEDR_{in}$ (mSv y^{-1}) were considered as input data, while $ELCR_{total}$ was chosen as output (target) parameter. Different types of machine learning algorithms have been developed to offer solutions to problems in different domains. Depending on the training method, these algorithms are divided into three types: supervised, unsupervised, and reinforcement learning (Kim, 2017; Shehab et al., 2022). However, in this study, supervised learning of regression learner in Matlab R2021a version 9.10.0 1602886 was engaged.

The regression learner trains regression models to predict data. It explores data, selects features, specifies validation schemes, train models, and assesses results. Automated training can be performed to search for the best regression model type, including linear regression models (LR), regression trees (RT), Gaussian process regression models (GPR), support vector machines (SVM), ensembles of regression trees (ET), and neural network regression models (NN). These regression model types exhibit higher efficiency, strong precision, and fast estimation (Mohana, 2020; Yeh, 1998; Ziolkowski and Niedostatkiewicz, 2019).

The model outputs were generated by training seven input variables (^{226}Ra, ^{232}Th, ^{40}K, $AGDR_{ex}$, $AGDR_{in}$, $AEDR_{ext}$, and $AEDR_{in}$) and responses ($ELCR_{total}$ data) using SVM, NN, ET, LR, RT, and GPR in regression learner of machine learning algorithms. This study compared the adopted MLAs and determined the algorithm with the best performance metrics.

Training and testing data are distinct, and the significant issue for the process of machine learning is attaining a good fit for training and testing parameters. Thus, obtaining accurate data representing actual data for model testing is essential (Thilakarathna et al., 2020). As a result, Kim (2017) posited that a model developed from MLA needs to be generalized without overfitting. Therefore, data validation is necessary to prevent data overfitting. In achieving this, a 10-fold cross-validation method is usually used to validate and enhance the performance of the developed model (Feng et al., 2020; Pereira and Borysov, 2019; Yeh, 1998). This technique minimizes the bias related to the random sampling of the training dataset. Another significant aspect of 10-fold cross-validation is that it equally segments the actual datasets into 10 subsets, enhances the learning with 9 subsets, and validates the model with 1 subset (Ahmad et al., 2022; Kohavi, 1995; Song et al., 2021). Hence, an average accuracy of 10 times is obtained after repeating the operation 10 times successfully. Kohavi (1995) established that the 10-fold cross-validation method represents the generalization and credibility of the model performance. Therefore, this study applied 10-fold cross-validation owing to its excellent characteristics performance as earlier stated above. In addition, by default, holdout validation was set at 25 %.

4.1. Gaussian process regression (GPR)

Gaussian Process Regression is an algorithm that generates an output based on the input variables. It generates models for complex datasets (Thilakarathna et al., 2020). In this study, input and target variables were trained based on the operational concepts of squared exponential, matern 5/2, exponential, and rational squared by GPR, leading to predicted outcomes. The output results were compared and the best performance metrics were included for comparison to other MLAs. The regression of the Gaussian process produces function distributions as in Eq. (8) using mean (m) and covariance (c) functions:

$$f \sim GPR(m, c) \tag{8}$$

4.2. Linear regression (LR)

Linear regression is the simpler and faster prediction technique for large data sets to determine the causal effects of dependent factors on independent factors (Gucluer et al., 2021). The dependent and independent parameters are considered continuous and response parameters. This study applied all four LR techniques: linear, interactions, robust, and stepwise to maximize their benefits in predicting ELCR of agricultural by-products. Depending on the impact of the variables on the result, the prediction output is obtained as a linear combination of input variables multiplied by a scaling factor (Gucluer et al., 2021). Furthermore, an additional degree of freedom is provided by adding the bias coefficient (intercept) as an additional variable. The general form of the linear regression algorithm is given in Eq. (9):

$$y = a_o + a_1 b_1 + a_2 b_2 + a_3 b_3 + \ldots a_i b_i + \ldots a_n b_n \tag{9}$$

where y is the predicted output, a_o is the bias coefficient, a_1, a_2, a_3..., an are the scale factors, and b_1, b_2, b_3,..., x_n are the input variables.

4.3. Neural network (NN)

The neural network is a supervised learning tool using neurons that predicts problems using weights and biases (Thilakarathna et al., 2020; Yang et al., 2004). It has excellent computational skills and a high tolerance for incorrect data (Yang et al., 2004). In maximizing NN potentials, this study looked at all five major networks of NN: narrow, medium, wide, two-layer, and three-layer neural networks, using a rectified linear units (ReLU) activation operation with an iteration limit of 1000. In contrast to the sigmoid function, the ReLU function (R), as in Eq. (10) is computationally more efficient to calculate since only a maximum number needs to be chosen (Nair and Hinton, 2010). It does not perform expensive exponential operations, blow-up activation and vanishing gradients compared to Sigmoid. In practice, networks using ReLU show better convergence performance than Sigmoid (Nair and Hinton, 2010).

$$R = f(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases} \tag{10}$$

An artificial neural network comprises six main components: the activation function $f(\Sigma)$, the sum function (Σ), inputs (p_i), weights (w_{ij}), biases (b), and output (A). The NN generates the best fit model using the weight, bias, and activation function and compares the result with the actual output (Gucluer et al., 2021; Gunoglu et al., 2013; Saridemir et al., 2009; Serin et al., 2011; Thilakarathna et al., 2020).

4.4. Ensemble trees (ET)

The ensemble tree has often been used as an alternative technique to evaluate the output when the linear approach does not provide accurate results (Gucluer et al., 2021). These techniques improve the weak algorithm by using an infinite number of decision trees on the input data during the training phase (Ahmad et al., 2022), but are often used for regression trees (Derousseau et al., 2019). Ensemble trees show high predicted performance on extremely nonlinear datasets. In addition, ensemble trees produce a successful result from unequal problems of predicted parameters. However, these techniques have low intuitiveness in relation to the trained data (Derousseau et al., 2019). The use of ensemble trees is common in civil engineering to model concrete properties. Based on their specifics, this research trained and tested all data on boosting and bagging ensemble trees to increase their benefits.

4.5. Support vector method (SVM)

Support Vector Machine is a supervised learning technique that solves regression and classification problems. It is a non-parametric algorithm that uses kernel functions for its operations (Gucluer et al., 2021; Thilakarathna et al., 2020). The Support Vector Machine algorithm uses a linear technique to provide predictive outputs for a given input (Narayan, 2020). Another benefit of SVM is that it predicts the output variable based on the input data. It trains and tests data efficiently (Gucluer et al., 2021). Because of their distinctness, the study trained and tested all input and target variables on six mathematical concepts of the SVM, namely linear, quadratic, cubic, fine-Gaussian, mean-Gaussian, and coarse-Gaussian.

4.6. Regression trees (RT)

The regression trees model the target data based on the partitions among the input parameters by creating partitions in the predicted variables. The regression tree technique is simpler and easier to interpret. Also, it provides robustness in managing missing and rich data. The regression approach was considered for this study because it is simpler and easier to interpret (Gucluer et al., 2021; Salehi and Burgueno, 2018). Furthermore, the regression method applies specific techniques that learn from the input parameters and provides highly accurate results for the output variables, giving a classic advantage over the traditional regression approach (Salehi and Burgueno, 2018). This study maximizes the potential of RT by training and testing all input and target data on fine, medium and coarse techniques. As in Eq. (11) the regression tree reduces the cost of complex functions (c) by scaling down an additional node selected for the model (Derousseau et al., 2019).

$$c = t(E) + \beta \times n(R) \tag{11}$$

Where $t(E)$ signifies training error, β is the regulator determined from cross-validation, and $n(R)$ is the number of regression tree leaves.

5. Performance indicators

The performance counters, also known as error metrics, measure the performance rate of each algorithm. Therefore, regardless of the machine learning technique used, it is important to determine the error metrics between the true and the predicted data. Therefore, this study examined four different performance indicators: the regression coefficient of determination (R^2), the mean absolute error (MAE), the mean square error (MSE), and the root mean square error (RMSE). The higher R^2, the stronger the prediction. However, the lower the MAE, MSE, and RMSE, the better the prediction (Willmott and Matsuura, 2005). The performance indicators are given in Eqs. (14)–(17):

Table 1
Radiological properties of agricultural byproducts examined.

Material	S (Bq kg^{-1})			AGDR (nGy h^{-1})		AEDR (mSv y^{-1})		ELCR			Ref.
	^{226}Ra	^{232}Th	^{40}K	ex	in	ex	in	out	in	total	
RHA	6	16	505	33	47	0.041	0.32	0.00014	0.00113	0.00127	Sas et al. (2019)
Average	6	16	505	33	47	0.041	0.32	0.00014	0.00113	0.00127	
MS	14.1	8.5	137	17	24	0.021	0.17	0.00007	0.00058	0.00066	Alam et al. (1999)
MS	9.9	6.1	96.2	12	17	0.015	0.12	0.00005	0.00041	0.00047	Alam et al. (1999)
MS	9.8	5.7	69.8	11	15	0.013	0.10	0.00005	0.00037	0.00041	Alam et al. (1999)
MS	12.2	5.6	110	14	19	0.017	0.13	0.00006	0.00046	0.00052	Alam et al. (1999)
MS	8.8	5.3	105	12	16	0.014	0.11	0.00005	0.00039	0.00044	Alam et al. (1999)
MS	8.4	4.4	81.4	10	14	0.012	0.10	0.00004	0.00033	0.00038	Alam et al. (1999)
MS	0.5	1.3	198	9	13	0.011	0.09	0.00004	0.00031	0.00035	Krmpotic et al. (2015)
MS	8.5	4.7	389	23	32	0.028	0.22	0.00010	0.00077	0.00087	Krmpotic et al. (2015)
MS	2.0	2.3	188	10	14	0.012	0.10	0.00004	0.00034	0.00039	Krmpotic et al. (2015)
MS	5.7	6.9	377	23	32	0.028	0.22	0.00010	0.00076	0.00086	Krmpotic et al. (2015)
Average	7.99	5.08	175	14	20	0.017	0.14	0.00006	0.00047	0.00054	
POC	6.89	4.46	571	29	41	0.037	0.20	0.00013	0.0001	0.00113	Karim et al. (2018)
POC	6.90	4.63	571	29	41	0.037	0.20	0.00013	0.0010	0.00113	Karim et al. (2018)
POC	5.85	4.40	573	29	40	0.036	0.19	0.00013	0.00098	0.00111	Karim et al. (2018)
POC	7.01	4.87	599	31	43	0.038	0.30	0.00013	0.00105	0.00118	Karim et al. (2018)
POC	5.56	4.21	604	30	42	0.037	0.29	0.00013	0.00102	0.00115	Karim et al. (2018)
POC	6.23	4.76	587	30	42	0.037	0.29	0.00013	0.00102	0.00115	Karim et al. (2018)
Average	6.41	4.56	584	30	42	0.04	0.25	0.00013	0.00101	0.00114	
POFA	8.16	6.14	441	26	36	0.032	0.25	0.00011	0.00087	0.00098	Karim et al. (2018)
POFA	7.74	6.41	450	26	36	0.032	0.25	0.00011	0.00088	0.00100	Karim et al. (2018)
POFA	7.95	7.13	457	27	38	0.033	0.26	0.00012	0.00091	0.00103	Karim et al. (2018)
POFA	8.75	7.98	421	26	37	0.032	0.25	0.00011	0.00089	0.00100	Karim et al. (2018)
POFA	9.03	7.56	498	29	41	0.036	0.28	0.00013	0.00099	0.00112	Karim et al. (2018)
POFA	6.98	6.91	413	24	34	0.030	0.24	0.00011	0.00083	0.00093	Karim et al. (2018)
Average	8.10	7.02	447	26	37	0.030	0.26	0.00012	0.00090	0.00101	
Global	33	45	420	59	84	0.07	0.41	0.00029	0.00116	0.00145	UNSCEAR (2000, 2008)

UNSCEAR: United Nations Scientific Committee on the Effects of Atomic Radiation (World Population-Weighted Average Value).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i^{pred} - y_i^{true})^2}{\sum_{i=1}^{n}(y_i^{pred} - \bar{y}^{true})^2} \tag{14}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(y_i^{pred} - y_i^{true}) \tag{15}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i^{pred} - y_i^{true})^2 \tag{16}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{pred} - y_i^{true})^2} \tag{17}$$

6. Results and discussion

6.1. Agricultural byproducts

Table 1 shows the radiological properties of the investigated agricultural by-products. The results showed that the specific activities of the agricultural by-products studied met the UNSCEAR (2000, 2008) recommended world population weighted averages, with the exception of RHA, POC and POFA, which were ^{40}K above the recommendation of 420 Bq kg^{-1}. This higher level of ^{40}K isotopes in RHA, POC and RHA could be related to the geological sources, treatment methods, processing patterns and production techniques of the by-products (Beretka and Mathew, 1985; Kovler, 2012; Sas et al., 2017, 2019). However, considering the global range, the named by-products (RHA, POC and POFA) met the recommended values of 140–850 Bq kg^{-1} for each building material (UNSCEAR, 2008). Also compared to the EC (1999) and Nuclear Energy Agency-Organization for Economic Co-operation and Development (NEA-OECD) (1979) recommendations, the specific activities of all investigated agricultural by-products met the mean values of 50, 50 and 500 Bq kg^{-1} for ^{226}Ra, ^{232}Th, and ^{40}K. Likewise, the results obtained in this study are consistent with the IAEA (2014) recommending the specific activities of all building materials in the range of 100–600, 30–300 and 100–1200 Bq kg^{-1} for ^{226}Ra, ^{232}Th, and ^{40}K isotopes.

Table 1 shows that the results of the external and internal absorbed gamma dose rates of RHA, MS, POC and POFA were lower than the world averages of 59 and 84 nGy h^{-1}. Similarly, the external and internal annual effective dose rates of all agricultural by-products studied were lower than the world averages of 0.071 and 0.41 mSv y^{-1}, indicating a possible suitability of the by-products considered.

The mean value of lifetime cancer excess risk, as reported in Table 1, was 0.14×10^{-3} outdoors, 1.13×10^{-3} indoors and 1.27×10^{-3}

Table 2
Performance indicators of trained and tested MLAs.

	Training/validation results				Test results			
MLA	RMSE	R^2	MSE	MAE	RMSE	R^2	MSE	MAE
LR	0.0000047	1.00	0.0000000	0.0000038	0.0000026	1.00	0.00000000	0.0000022
SVM	0.0000438	0.98	0.0000000	0.0000359	0.0000532	0.97	0.00000000	0.0000468
RT	0.0000806	0.94	0.0000000	0.0000062	0.0000707	0.95	0.00000000	0.0000549
ET	0.0001135	0.88	0.0000000	0.0000940	0.0000754	0.94	0.00000000	0.0000599
GPR	0.0000000	0.00	0.0000001	0.0002875	0.0003091	0.00	0.00000000	0.0002759
NN	0.0138610	-1822	0.0001921	0.0098816	0.0009975	-9.41	0.00000100	0.0007546

Table 3
Previous studies on performance metrics of MLAs.

Ref.	Metric	Training/MLAs						Testing/MLAs						Study
		LR	SVM	RT	NN	GPR	ET	LR	SVM	RT	NN	GPR	ET	
Gucluer et al. (2021)	R	0.8500	0.8400	0.8600	0.8600	–	–	–	–	–	–	–	–	ML techniques for estimating the compressive strength of concrete
	R^2	0.8500	0.8400	0.8600	0.8600	–	–	–	–	–	–	–	–	
	MAE	2.9000	2.9300	2.5900	3.4200	–	–	–	–	–	–	–	–	
	RMSE	4.0200	4.0300	3.7700	4.4800	–	–		–	–	–	–	–	
Nguyen et al. (2020a, 2020b)	R	–	–	–	0.9930	–	–	–	–	–	0.985	–	–	Analyzing the compressive strength of geopolymer concrete using ML approaches
	MAE	–	–	–	0.8420	–	–	–	–	–	1.115	–	–	
	RMSE	–	–	–	1.4740	–	–	–	–	–	1.824	–	–	
Behnood and Golafshani (2020)	R	–	–	0.954	–	–	–	–	–	0.864	–	–	–	ML study of compressive strength of concrete containing waste foundry sand
	R^2	–	–	0.945	–	–	–	–	–	0.932	–	–	–	
	MAE	–	–	2.985	–	–	–	–	–	3.427	–	–	–	
	RMSE	–	–	4.312	–	–	–	–	–	4.715	–	–	–	
Liu et al. (2022)	R^2	–	0.8709	0.9810	0.9274	–	–	–	0.8709	0.9810	0.9274	–	–	ML techniques in predicting the allegator and longitudinal cracks by improving the asphalt mix design
	MAE	–	2.3200	2.1600	1.7800	–	–	–	–	–	–	–	–	
	MSE	–	41.970	29.560	24.980	–	–	–	–	–	–	–	–	
	R^2	–	0.8556	0.9530	0.8791	–	–	–	0.7162	0.7775	0.7334	–	–	
	MAE	–	364.57	308.12	330.91	–	–	–	–	–	–	–	–	
	MSE	–	507.96	398.32	477.24	–	–	–	–	–	–	–	–	
Duan et al. (2013)	R^2	–	–	–	0.9970	–	–	–	–	–	0.9914	–	–	ANNs for predicting the MoE of recycled aggregate-concrete
	RMSE	–	–	–	1.3455	–	–	–	–	–	2.4223	–	–	
Paixao et al. (2022)	R^2	–	0.7900	–	0.8200	0.8200	–	–	0.3700	–	0.5100	0.5900	–	ML techniques to predict the compressive strength of concrete
	MAE	–	2.2600	–	2.2600	1.9600	–	–	4.0400	–	3.2300	3.0400	–	
	RMSE	–	3.7300	–	3.4000	3.4300	–	–	4.6700	–	4.0900	3.7500	–	
Khatti and Grover (2021)	R	0.9744	0.9539	0.9381	0.9780	–	–	0.9379	0.9273	0.9565	0.9090	–	–	Computation of compressive strength of GGBS-concrete using ML
	MAE	1.7274	2.2314	2.1103	0.0200	–	–	2.6946	2.5087	1.2288	2.7923	–	–	
	RMSE	2.2053	3.0271	3.4227	0.0814	–	–	3.3224	3.4505	1.8514	3.4055	–	–	
Chopra et al. (2018)	R^2	–	–	0.8604	0.9769	–	–	–	–	0.8008	0.9500	–	–	ML techniques in predicting concrete compressive strength
	RMSE	–	–	2.0111	0.7176	–	–	–	–	2.2514	0.9100	–	–	
Dabiri et al. (2022)	R^2	0.8000	–	–	0.9801	–	–	–	–	–	–	–	–	ML techniques in compressive strength of concrete with recycled aggregate
	MAE	8.6800	–	–	2.0900	–	–	–	–	–	–	–	–	
	RMSE	11.170			2.9400	–	–	–	–	–	–	–	–	
Ahmad et al. (2022)	R^2	–	–	0.9033	–	–	0.9683	–	–	0.8900	–	–	0.8300	Analyzing the compressive strength of geopolymer concrete using ML approaches
	MAE	–	–	2.6200	–	–	1.5100	–	–	1.4800	–	–	2.2300	
	MSE	–	–	11.400	–	–	3.7500	–	–	2.3000	–	–	5.1600	
	RMSE	–	–	3.3800	–	–	1.9400	–	–	1.5200	–	–	2.2700	
Naseri et al. (2020)	R	0.9116	0.9000	–	0.9849	–	–	0.6842	0.8990	–	0.9530	–	–	Designing sustainable concrete mixture by developing a new machine learning technique
	R^2	0.8311	0.8099	–	0.9701	–	–	0.4682	0.8082	–	0.9081	–	–	
	MAE	5.6216	5.9155	–	2.0424	–	–	8.4486	5.5761	–	2.8937	–	–	
	MSE	43.448	49.238	–	8.9642	–	–	208.15	45.283	–	18.995	–	–	
	RMSE	6.5915	7.0170	–	2.9940	–	–	14.427	6.7293	–	4.3583	–	–	
Salami et al. (2021)	R^2	–	0.9820	–	–	–	–	–	0.9540	–	–	–	–	Predicting ternary-blend concrete strength using MLAs
	RMSE	–	2.2300	–	–	–	–	–	3.3500	–	–	–	–	
Nyguyen et al. (2020a, 2002b)	R	–	0.9500	–	–	–	–	–	0.9600	–	–	–	–	Efficient machine learning models for prediction of concrete strengths
	MAE		3.7900	–	–	–	–	–	0.2700	–	–	–	–	
	RMSE	–	5.0000	–	–	–	–	–	0.3900	–	2.4223	–	–	
Shamsabadi et al. (2022)	R^2	0.4400	0.9100	–	0.9700	–	–	–	–	–	–	–	–	ML-based compressive strength modeling of concrete incorporating waste marble powder
	MAE	7.7100	2.4800	–	1.3500	–	–	–	–	–	–	–	–	
	MSE	92.490	17.450	–	5.2600	–	–	–	–	–	–	–	–	
	RMSE	9.6200	4.1800	–	2.2900	–	–	–	–	–	–	–	–	
Feng et al. (2020)	R^2	–	0.9960	–	0.9970	–	–	–	0.9790	–	0.9810	–	–	ML-based compressive strength prediction for concrete: Adaptive boosting
	MAE	–	1.2600	–	1.2600	–	–	–	1.660	–	1.6600	–	–	
	RMSE	–	1.5500	–	1.5300	–	–	–	2.2600	–	2.2500	–	–	

(*continued on next page*)

Table 3 (*continued*)

Ref.	Metric	Training/MLAs						Testing/MLAs						Study
		LR	SVM	RT	NN	GPR	ET	LR	SVM	RT	NN	GPR	ET	
Song et al. (2021)	R^2	–	–	0.7510	0.8124	–	0.9571	–	–	0.7600	0.7700	–	0.8700	Predicting the compressive strength of concrete with fly ash admixture using MLAs
	MAE	–	–	4.5500	4.4800	–	3.6900	–	–	5.9900	7.4700	–	6.7800	
	MSE	–	–	32.660	28.740	–	24.760	–	–	9.5900	7.6900	–	6.7700	
	RMSE	–	–	5.3600	5.3600	–	4.9700	–	–	3.1000	2.7700	–	2.600	
Thilakarathna et al. (2020)	R^2	0.7700	0.8300	0.8200	0.9700	0.8900	–	–	–	–	–	–	–	MLAs in embodied CO_2 and emissions of concrete
	MAE	7.300	6.300	6.3000	–	4.800	–	–	–	–	–	–	–	

overall for RHA. MS showed 0.06 10^{-3} outdoors, 0.47×10^{-3} indoors and 0.54×10^{-3} overall. Additionally, POC was 0.13×10^{-3} outdoors, 1.01×10^{-3} indoors, and 1.14×10^{-3} overall. And POFA returned 0.12×10^{-3} as Outdoor, 0.90×10^{-3} as Indoor, and 1.01×10^{-3} overall. In comparison, these results were below the world population-weighted averages reported by UNSCEAR (2000, 2008), which were 0.29×10^{-3} for outdoor, 1.16×10^{-3} for indoor and 1.45×10^{-3} for the overall excess lifetime cancer risk. In terms of sequential lifetime factors, therefore, MS shows the best suitability, followed by POFA, POC and RHA.

Most importantly, as shown in Table 1, the results showed that MS, POFA, POC and RHA do not pose a cancer risk. Therefore, MS, POFA, POC and RHA can be used as building and construction materials, but with caution.

6.2. Machine learning algorithms

Given the zero coefficient of determination (R^2), GPR, and NN obviously have not learned anything. Therefore, Table 2 shows the performance indicators (R^2, MAE, MSE and RMSE) for the four different selected model types. As mentioned above, it is important to note that the radiological parameters of the agricultural by-products studied were all trained, validated, and tested in twenty-four different models from six different machine learning algorithms of NN, ET, GPR, LR, RT and SVM. The input parameters of the radiological parameters were ^{226}Ra, ^{232}Th, ^{40}K, $AGDR_{ex}$, $AGDR_{in}$, $AEDR_{in}$, and $AEDR_{ex}$, while $ELCR_{total}$ was chosen as the target variable to predict the excess lifetime cancer risks. Table 2 shows that the linear regression algorithm had the best performance indicators for modeling the excess lifetime cancer risks of agricultural by-products based on the training and testing scores of 1.00 for R^2 and the lowest mean square error values of 0.0000047 for training scores and 0.0000026 for test results.

In general, the mean absolute error (MAE) shows the weight and importance of the error due to the different values and error scales (Willmott and Matsuura, 2005). Therefore, the error-efficient performance of each model is based on the mean absolute error. The MAE in training and testing the datasets, as displayed in Table 2, was far from zero, indicating a good correlation. However, the MAE of LR technique in training the ELCR datasets of agricultural by-products was 89.42 %, 38.71 %, 95.96 %, 98.86 %, and 99.96 % lower than the SVM, RT, ET, GPR, and NN techniques, respectively. Similarly, the MAE in testing the ELCR datasets of agricultural by-products using the LR technique was 95.30 %, 95.99 %, 96.33 %, 99.71 %, and 99.20 % lower than SVM, RT, ET, NN, and GPR methods, respectively. The evidence from these results indicates that there is a linear correlation between the specific activities and the radiological hazards of agricultural by-products. These agree with the representations in Eqs. (1)–(7). Thus, the best correlation between the true and predicted agricultural by-product variables exists in the LR algorithm. A machine learning algorithm application to the excess lifetime cancer risks of agricultural by-products has not yet been studied. Nevertheless, Table 3 contains the performance indicators for the various model types used in this study that are also used in other disciplines to support the findings. Ultimately, these results indicated that the LR algorithm could optimize the radiological indices of agricultural by-products to achieve the expected ELCR output.

It can be seen from Table 2 that the performance indicators of MLAs for predicting ELCR of agricultural by-products are classified in order of LR (linear), SVM (linear), RT (fine), and ET (boosted). As previously explained, the higher the R^2 to one, the better the model predicts the data. The lower the RMSE, MAE and MSE to zero, the better the model performs. As a result, the LR model is given in Eq. (18):

$$ELCR_{total} = 0.0000037965 + 0.000013634 \; (^{226}Ra) + 0.000017709 \; (^{232}Th) + 0.000001232 \; (^{40}K) - 0.00000024792 \; (AGDR_{out}) + 0.000001341 \; (AGDR_{in}) + 0.053105 \; (AEDR_{out}) + 0.000021452 \; (AEDR_{in}) \quad (18)$$

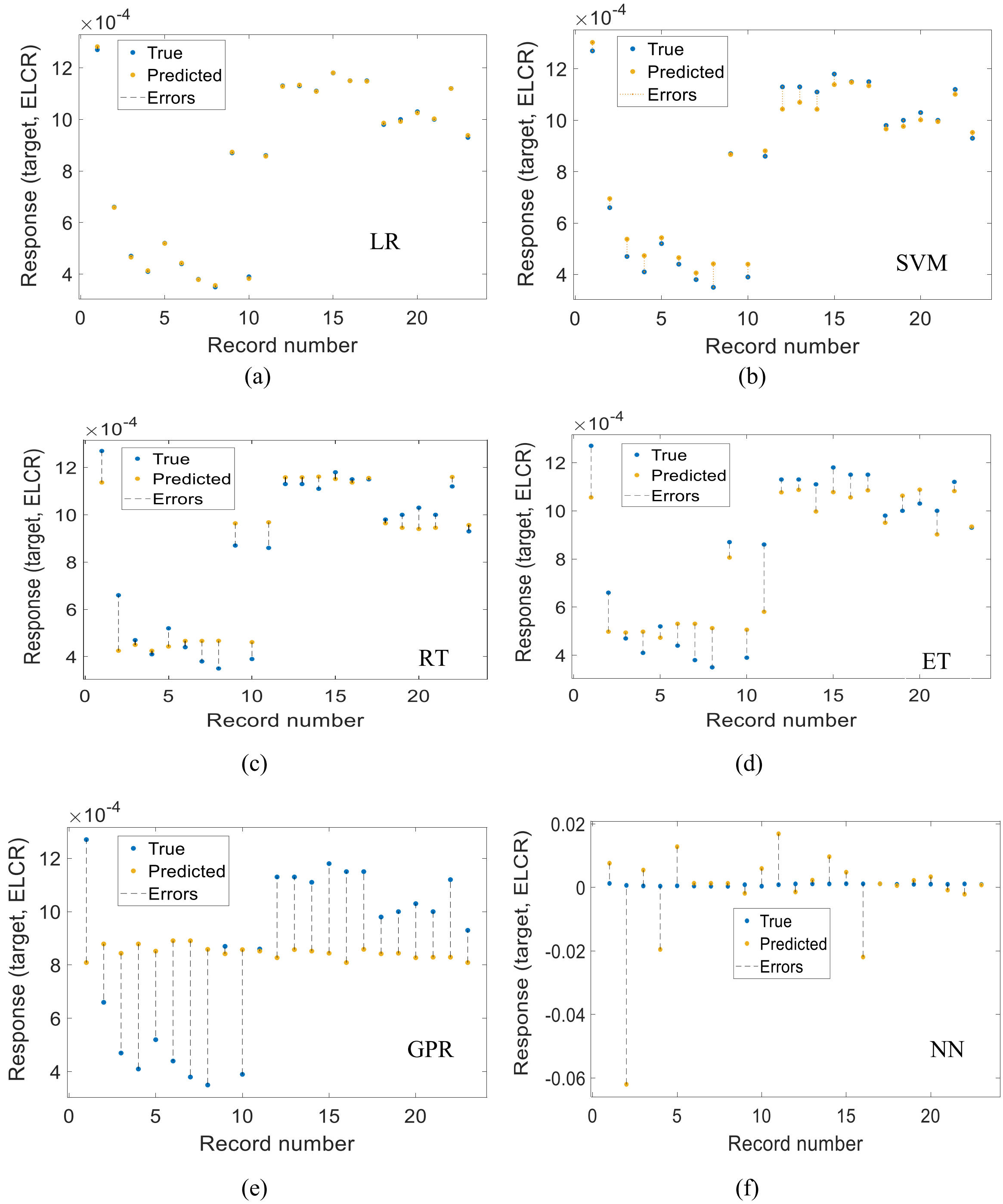

Fig. 1. Response plots for machine learning algorithms adopted.

A systematic understanding for predicting excess lifetime cancer risk from agricultural by-products via MLAs is still lacking. However, the accuracy and validity of the performance indicators obtained in this study are confirmed by comparing them (Table 2) to those from previous studies (Table 3) that used similar MLAs but different materials and input hyper-variables. The results are summarized in Table 3. As shown in Table 3, it is now well established through comparisons from a variety of studies that the MLAs adopted demonstrated superior robustness and performance with the lowest prediction errors.

6.2.1. Response plots

Fig. 1 shows the response plots for $ECLR_{total}$ and record number to visualize the relationship between the input parameters and the response ($ECLR_{total}$). Each radiological index included 23 observations, a total of 161 observations for input parameters (^{226}Ra, ^{232}Th, ^{40}K, $AGDR_{ex}$, $AGDR_{in}$, $AEDR_{in}$, and $AEDR_{ex}$) and 23 observations for the target (response) variable ($ELCR_{total}$). Each prediction was obtained based on the trained model. The LR algorithm, as shown in Fig. 1, showed a strong correlative response with little or no error. Because LR assumes a linear relationship between the input variable and the output variable (target), which is absolutely linear in the true sense of the word. Some data sets in the SVM algorithm showed a number of scatters with error traces. Also, RT, ET, GPR, and NN techniques showed a range of scatter with a larger margin of error. These results validate the performance indicators presented in Table 2 for all MLAs used.

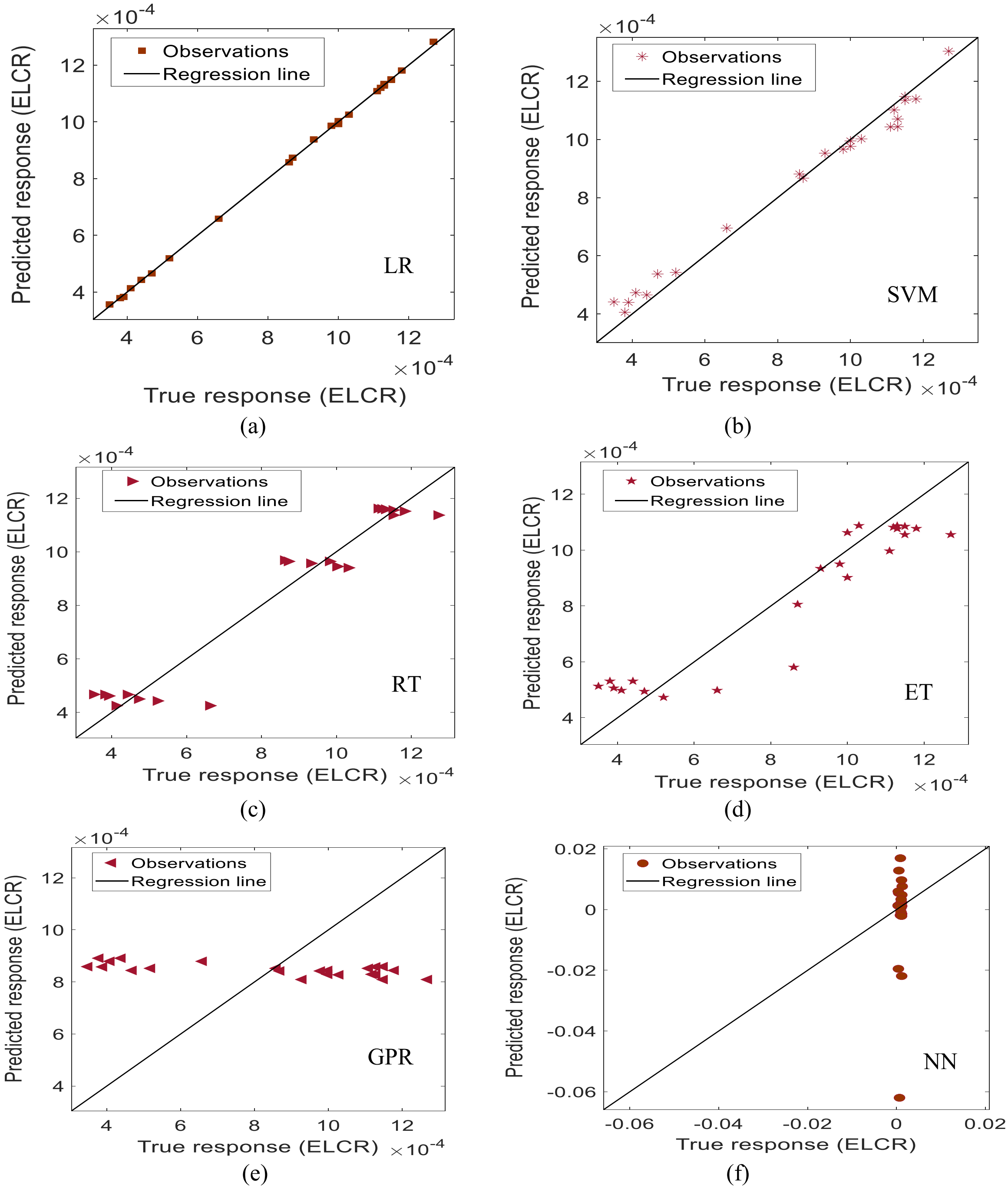

Fig. 2. Predicted versus true responses machine learning algorithms used.

6.2.2. Predicted and true responses

Fig. 2 shows the predicted versus true ELCR for the trained data. Results showed perfect regression models for LR and SVM algorithms, such that all data points fell on diagonal (regression) lines. However, the RT and ET algorithms encountered few outliers, where data points were roughly symmetrically spread around the regression line. However, the GPR and NN algorithms showed an imperfect model as the data are far from the regression line. In support of these results, few or no studies have been conducted to compare predicting the ELCR of agricultural by-products using machine learning algorithms. However, it is noteworthy to note that these results confirm the previous studies (Ahmad et al., 2022; Gucluer et al., 2021; Nguyen et al., 2020a, 2020b; Thilakarathna et al., 2020) such that LR and SVM Algorithms gave a perfect prediction of concrete compressive strength, validating the efficiency of these algorithms.

6.2.3. Output performance

Fig. 3 shows the model outputs for all machine learning algorithms used. The result outputs established a strong and perfect prediction for LR and SVM algorithms because the data fit perfectly with the regression line. The relationship between the true response (ELCR) and the input arguments is linear. Besides, LR and SVM use linear and linear kernel functions, respectively, thus, the true responses for LR and SVM, as indicated in Fig. 2, are equal to the predicted responses, signifying a perfect correlation. Moreover, RT and ET algorithms also gave a perfect result due to their strong correlation (R) of about 0.97 and 0.98 respectively. However, GPR and NN showed strongly identifiable outliers, rendering the models imperfect. The reason for this poor prediction is that GPR models are non-parametric kernel-based probabilistic models, hence, the use of exponential kernel or Matern 5/2 functions does not

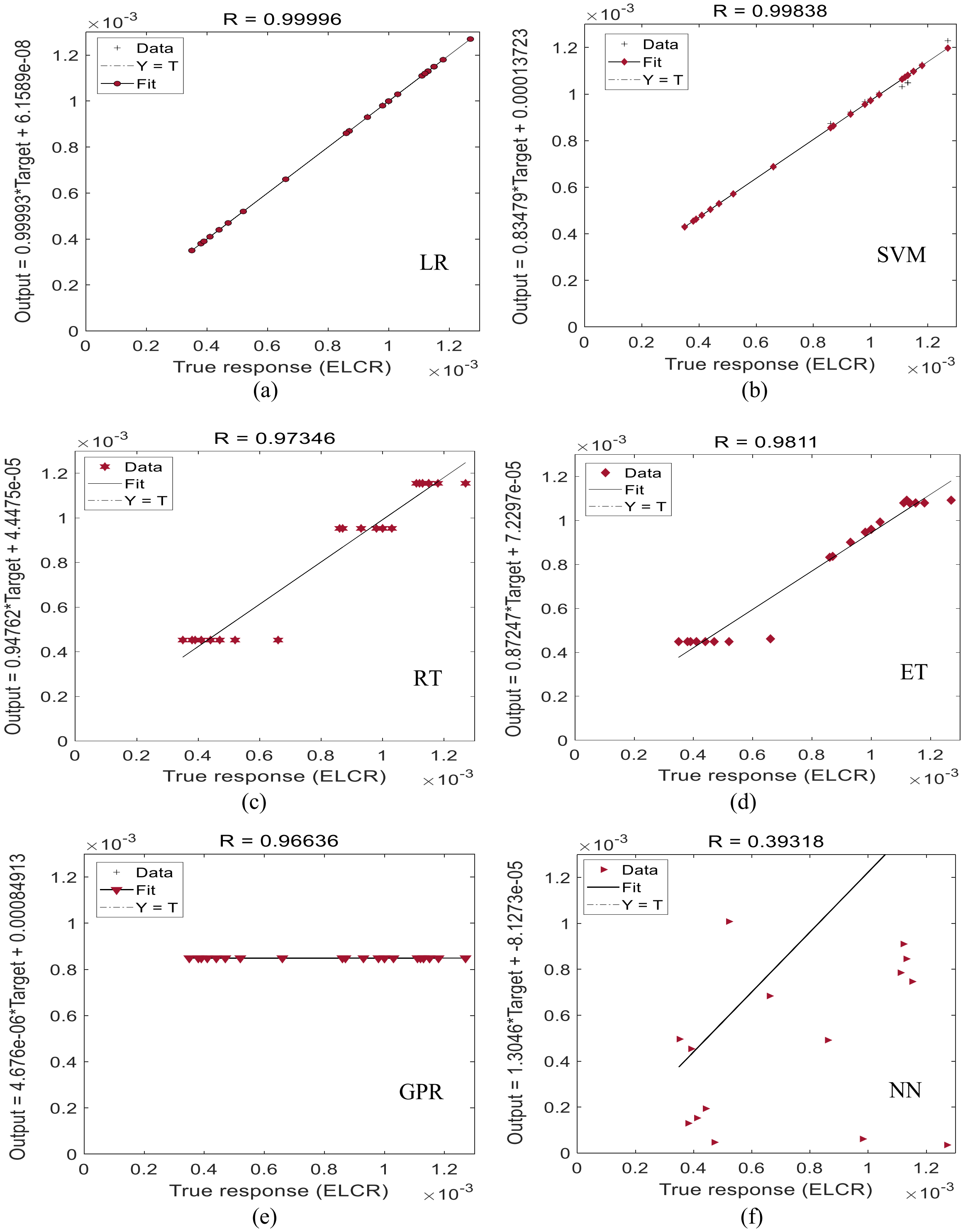

Fig. 3. Model outputs for machine learning algorithms engaged.

seem to be a good predictor of ELCR datasets. Likewise, neural network models are fully connected layers of various sizes, which seems not to be a good predictor of ELCR datasets. Sometimes NNs may not be able to find their minimal error value to give back an accurate output to the user. This happens when neurons have high weights so that the optimization processes like gradient descent will not be able to find global minima, resulting in inaccurate output (Cohen, 2021), as shown in Fig. 3(f) for NN. The results presented in Fig. 3 are consistent with performance indicators, response charts and regression charts highlighted in Table 2, Fig. 1 and Fig. 2. There are similarities between the output power expressed by LR and SVM algorithms in this study and those reported by Thilakarathna et al. (2020). In this context, data on the compressive strength of concretes after 28 days of curing were collected and trained with the various machine learning algorithms. After training, the model outputs gave R-values of 0.8503 and 0.8424 for LR and SVM, respectively. The LR approach ultimately offers the optimum fit for training and testing the ELCR datasets of agricultural by-products for better output performance.

7. Conclusions

This study examined the radionuclide content of agricultural by-products and predicted their excess lifetime cancer risks (ELCR) using six different machine learning algorithms. The main findings of this research revealed that the naturally occurring radionuclides in RHA, MS, POC, and POFA do not pose a cancer risk. However, the potential users should apply them with caution. The excess lifetime cancer risks in RHA, MS, POC, and POFA were approximately 12 %, 63 %, 22 %, and 31 % lower than the UNSCEAR world population-weighted averages. The LR represents the most accurate technique for predicting the ELCR of surveyed agricultural byproducts. In comparison to SVM, RT, ET, GPR, and NN algorithms, the LR method yielded the best error-free efficiency in terms of mean absolute error for training and testing the ELCR datasets.

Overall, the results help identify, understand and predict the cancer-related risks associated with the use of agricultural byproducts derived from recycled agricultural waste. In addition, it provides the reference data for establishing the radiation monitoring framework and recommendations for the use of agricultural byproducts containing NORs. However, for further studies, the range of the ELCR of the agricultural datasets should be considered to provide higher efficiency in predicting the ELCR of agricultural byproducts.

Ethical approval

This article has no human participants or animals performed by any of the authors.

Data Availability

Data used are included in the manuscript.

Declaration Interest Statement

The authors show credit to the sources in the manuscript. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The raw/processed data required to reproduce these findings cannot be shared at this time as the Data also forms part of an ongoing study. The authors declare that the manuscript is the authors' original work and has not been published before. The authors also declare that the article contains no libelous or unlawful statements and does not infringe on the rights of others.

Acknowledgement

The authors want to thank the Covenant University Centre for Research, Innovation and Development for the favorable environment and supporting this research.

References

Abdullahi, A., Ismail, A.F., Samat, S., 2019. Determination of indoor doses and excess lifetime cancer risks caused by building materials containing natural radionuclides in Malaysia. Nucl. Eng. Technol. 51, 325–336. https://doi.org/10.1016/j.net.2018.09.017

Aghajanian, A., Thomas, C., Sainz-Aja, J., 2022. The use of rice hush ash in eco-concrete. In: The structural integrity of recycled aggregate concrete produced with fillers and pozzolans. Elsevier, pp. 171–197. https://doi.org/10.1016/B978-0-12-824105-9.00006-8

Ahmad, A., Ahmad, W., Aslam, F., Joyklad, P., 2022. Compressive strength prediction of fly ash-based geopolymer concrete via advanced machine learning techniques. Case Stud. Constr. Mater. 16, e00840. https://doi.org/10.1016/j.cscm.2021.e00840

Andrew, R.M., 2018. Global CO_2 emissions from cement production, 1928–2017. Earth Syst. Sci. Data 10, 2213–2239. https://doi.org/10.5194/essd-10-2213-2018

Aprianti, E., Shafigh, P., Bahri, S., Farahani, J.N., 2015. Supplementary cementitious materials origin from agricultural wastes–a review. Constr. Build. Mater. 74, 176–187. https://doi.org/10.1016/j.conbuildmat.2014.10.010

Behnood, A., Golafshani, E.M., 2020. Machine learning study of the mechanical properties of concretes containing waste foundry sand. Constr. Build. Mater. 243, 118152. https://doi.org/10.1016/j.conbuildmat.2020.118152

Belaid, F., 2022. How does concrete and cement industry transformation contribute to mitigating climate change challenges? Res. Conserv. Recycl. Adv. 15, 200084. https://doi.org/10.1016/j.rcradv.2022.200084

Beretka, J., Mathew, P.J., 1985. Natural radioactivity of Australian building materials, industrial wastes and by products. Health Phys. 48, 87–95. https://doi.org/10.1097/00004032-198501000-00007

Buasri, A., Chaiyut, N., Loryuenyong, V., Worawanitchaphong, P., Trongyong, S., 2013. Calcium oxide derived from waste shells of mussel, cockle, and scallop as the heterogeneous catalyst for biodiesel production. Sci. World J. 460923, 1–7. https://doi.org/10.1155/2013/460923

Canadian Nuclear Safety Commission, 2012. Introduction to Radiation. Available at: 〈http://www.nuclearsafety.gc.ca/〉, (Accessed 31 May 2022).

CEU, 2014. Council Directive 2013/59/Euratom of 5 December 2013 laying down basic safety standards for protection against the dangers arising from exposure to ionizing radiation, and repealing directives 89/618/Euratom, 90/641/Euratom, 96/29/Euratom, 97/43/Euratom and 2003/122/Euratom. Off. J. Eur. U. N., vol. 13, pp. 1–73. Available at: 〈https://ec.europa.eu/energy/sites/ener/files/documents/CELEX-32013L0059-EN-TXT.pdf〉, (Accessed 20 August 2022).

Chinnu, S.N., Minnu, S.N., Bahurudeen, A., Senthilkumar, R., 2021. Reuse of industrial and agricultural by-products as pozzolan and aggregates in lightweight concrete. Constr. Build. Mater. 302, 124172. https://doi.org/10.1016/j.conbuildmat.2021.124172

Chopra, P., Sharma, R.K., Kumar, M., Chopra, T., 2018. Comparison of machine learning techniques for the prediction of compressive strength of concrete. Adv. Civ. Eng. 5481705, 1–9. https://doi.org/10.1155/2018/5481705

Cohen, S.M.D., 2021. The evolution of machine learning: past, present, and future. Artificial Intelligence and Deep Learning in Pathology. Elsevierhttps://doi.org/10.1016/B978-0-323-67538-3.00001-4

Dabiri, H., Kioumarsi, M., Kheyroddin, A., Kandiri, A., Sartipi, F., 2022. Compressive strength of concrete with recycled aggregate; a machine learning-based evaluation. Clean. Mater. 3, 100044. https://doi.org/10.1016/j.clema.2022.100044

Darby, S., Hill, D., Auvinen, A., Barros-Dios, J.M., Baysson, H., Bochicchio, F., Deo, H., Falk, R., Forastiere, F., Hakama, M., Heid, I., Kreienbrock, L., Kreuzer, M., Lagarde, F., Mäkeläinen, I., Muirhead, C., Oberaigner, W., Pershagen, G., Ruano-Ravina, A., Ruosteenoja, E., Rosario, A.S., Tirmarche, M., Tomáscaronek, L., Whitley, E., Wichmann, H.-E., Doll, R., 2005. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. BMJ 330 (7485), 223–230. https://doi.org/10.1136/bmj.38308.477650.63

Derousseau, M.A., Laftchiev, E., Kasprzyk, J.R., Rajagopalan, B., Srubar, W.V., 2019. A comparison of machine learning methods for predicting the compressive strength of field-placed concrete. Constr. Build. Mater. 228, 116661. https://doi.org/10.1016/j.conbuildmat.2019.08.042

Duan, Z.H., Kou, S.C., Poon, C.S., 2013. Using artificial neural networks for predicting the elastic modulus of recycled aggregate concrete. Constr. Build. Mater. 44, 524–532. https://doi.org/10.1016/j.conbuildmat.2013.02.064

EC, 1999. Radiological Protection Principles Concerning the Natural Radioactivity of Building Materials (Radiation Protection Report – RP-112). European Commission, Luxembourg. Available at: 〈https://ec.europa.eu/energy/sites/ener/files/documents/112.pdf〉, (Accessed 23 May 2022).

Faghihi, R., Mehdizadeh, S., Sina, S., 2011. Natural and artificial radioactivity distribution in soil of Fars province, Iran. Radiat. Prot. Dosim. 145, 66–74. https://doi.org/10.1093/rpd/ncq367

Feng, D.C., Liu, Z.T., Wang, X.D., Chen, Y., Chang, J.Q., Wei, D.F., Jiang, Z.M., 2020. Machine learning-based compressive strength prediction for concrete: an adaptive boosting approach. Constr. Build. Mater. 230, 117000. https://doi.org/10.1016/j.conbuildmat.2019.117000/

Ghalehnovi, M., Roshan, N., Taghizadeh, A., Shamsabadi, E.A., Hadigheh, S.A., de Brito, J., 2022. Production of environmentally friendly concrete incorporating bauxite residue and silica fume. J. Mater. Civ. Eng. 34 (2), 04021423. https://doi.org/10.1061/(ASCE)MT.1943-5533.0004060

Gucluer, K., Ozbeyaz, A., Goymen, S., Gunaydn, O., 2021. A comparative investigation using machine learning methods for concrete compressive strength estimation. Mater. Todays Commun. 27, 102278. https://doi.org/10.1016/j.mtcomm.2021.102278

Gunoglu, K., Demir, N., Akkurt, I., Demirci, Z.N., 2013. ANN modeling of the bremsstrahlung photon flux in tantalum target. Neural Comput. Appl. 23 (6), 1591–1595. https://doi.org/10.1007/s00521-012-1111-2

IAEA, 2014. Radiation Protection and Safety of Radiation Sources. International Basic Safety Standards, Vienna. Available at: 〈https://www-pub.iaea.org/MTCD/publications/PDF/Pub1578_web-57265295.pdf〉, (Accessed 21 May 2022).

Ibrahim, H.A., Abdul Razak, A., Abutaha, F., 2017. Strength and abrasion resistance of palm oil clinker pervious concrete under different curing method. Constr. Build. Mater. 147, 576–587. https://doi.org/10.1016/j.conbuildmat.2017.04.072

ICRP, 1990. Recommendations of the International Commission on Radiological Protection. vol. 21(No. 1–3), publication 60. 〈https://journals.sagepub.com/doi/pdf/10.1177/ANIB_21_1-3〉, (Accessed 3 July 2022).

ICRP, 1994. Protection against Rn-222 at Home and at Work. ICRP Publication 65. Annals of the ICRP. vol. 23, pp. 1–48. Available at: 〈https://journals.sagepub.com/doi/pdf/10.1177/ANIB_23_2〉, (Accessed 5 June 2022).

Imani, M., Adelikhah, M., Shahrokhi, A., Azimpour, G., Yadollahi, A., Kocsis, E., Toth-Bodrogi, E., Kovács, T., 2021. Natural radioactivity and radiological risks of common building materials used in Semnan Province dwellings, Iran. Environ. Sci. Pollut. Res. 28, 41492–41503. https://doi.org/10.1007/s11356-021-13469-6

Ishak, M.U.F., Othman, N.H.B., Sainudin, M.S., 2021. A review: effects of mussel shell ash as concrete mixture under sodium chloride exposure. Recent Trends Civ. Eng. Built

Environ. 2 (1), 824–833. ⟨http://penerbit.uthm.edu.my/periodicals/index.php/rtcebe⟩.
Joel, E.S., Maxwell, O., Adewoyin, O.O., Olawole, O.C., Arijaje, T.E., Embong, Z., Saeed, M.A., 2019. Investigation of natural environmental radioactivity concentration in soil of coastaline area of Ado-Odo/Ota Nigeria and its radiological implications. Sci. Rep. 4219. https://doi.org/10.1038/s41598-019-40884-0
Karim, M.R., Khandaker, M.U., Asaduzzaman, K.H., Razak, H.A., Yusoff, S.B., 2018. Radiological risks assessment of building materials ingredients: palm oil clinker and fuel ash. Indoor Built Environ. 28 (4), 479–491. https://doi.org/10.1177/1420326X18776705
Khatti, S.W., Grover, K.S., 2021. Computation of compressive strength of ggbs mixed concrete using machine learning. Int. J. Recent Technol. Eng. 10 (4), 241–250. https://doi.org/10.35940/ijrte.D6631.1110421
Kim, P., 2017. MATLAB Deep Learning with Machine Learning, Neural Networks and Artificial Intelligence. Springer Book: Apress Berkeley, CA. Softcover ISBN 978-1-4842-2844-9.
Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. Int. Jt. Confr. Art. Int. 14, 1137–1145 http://ai.stanford.edu/~ronnyk/accEst.pdf.
Kovler, K., 2012. Radioactive materials. Toxicity of Building Materials. Woodhead Publishing Limited: Elsevier, pp 196–240. ⟨https://doi.org/10.1533/9780857096357.196⟩.
Krmpotic, M., Rozmaric, M., Barisic, D., 2015. Mussels (Mytilus galloprovincialis) as a bio-indicator species in radioactivity monitoring of Eastern Adriatic coastal waters. J. Environ. Radioact. 144, 47–51. https://doi.org/10.1016/j.jenvrad.2015.02.027
Liu, J., Liu, F., Gong, H., Fanijo, E.O., Wang, L., 2022. Improving asphalt mix design by predicting alligator cracking and longitudinal cracking based on machine learning and dimensionality reduction techniques. Constr. Build. Mater. 354, 129162. https://doi.org/10.1016/j.conbuildmat.2022.129162
Maxwell, O., Wagiran, H., Ibrahim, N., Lee, S.K., Embong, Z., Ugwuoke, P.E., 2015. Natural radioactivity and geological influence on subsurface layers at Kubwa and Gosa area of Abuja, Northcentral Nigeria. J. Radioanal. Nucl. Chem. 303, 821–830. https://doi.org/10.1007/s10967-014-3442-1
Mohana, M.H., 2020. The determination of ground granulated concrete compressive strength-based machine learning models. Period. Eng. Nat. Sci. 8 (2), 1011–1023. https://doi.org/10.21533/pen.v8i2.1344.g584
Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In: Pro. 27th Int. Conf. Machine Learning (ICML-10), pp. 807–14. ⟨https://dl.acm.org/doi/10.5555/3104322.3104425⟩.
Naseri, H., Jahanbakhsh, H., Hosseini, P., Nejad, F.M., 2020. Designing sustainable concrete mixture by developing a new machine learning technique. J. Clean. Prod. 258, 120578. https://doi.org/10.1016/j.jclepro.2020.120578
Nguyen, H., Vu, T., Vo, T.P., Thai, H.T., 2020a. Efficient machine learning models for prediction of concrete strengths. Constr. Build. Mater. 266, 120950. https://doi.org/10.1016/j.conbuildmat.2020.120950
Nguyen, K.T., Nguyen, Q.D., Le, T.A., Shin, J., Lee, K., 2020b. Analyzing the compressive strength of green fly ash based geopolymer concrete using experiment and machine learning approaches. Constr. Build. Mater. 247, 118581. https://doi.org/10.1016/j.conbuildmat.2020.118581
Nuclear Energy Agency-Organization for Economic Co-operation and Development (NEA-OECD), 1979. Exposure to radiation from radioactivity in building materials.
Okan, I., Yildiz, H., Kisbet, T., Erturk, S.M., Onder, H., 2022. Classification of retinoblastoma-1 gene mutation with machine learning-based models in bladder cancer. Heliyon 8 (4), e09311. https://doi.org/10.1016/j.heliyon.2022.e09311
Olthof, A.W., Shouche, P., Fennema, E.M., Ijpma, F.F.A., Koolstra, R.H.C., Stirler, V.M.A., van Ooijen, P.M.A., Cornelissen, L.J., 2021. Machine learning based natural language processing of radiology reports in orthopaedic trauma. Comput. Methods Prog. Biomed. 208, 16304. https://doi.org/10.1016/j.cmpb.2021.106304
Paixao, R.C.F., Penido, R.E., Cury, A.A., Mendes, J.C., 2022. Comparison of machine learning techniques to predict the compressive strength of concrete and considerations on model generalization. Rev. IBRACON Estrut. Mater. 15 (5), 15503. https://doi.org/10.1590/S1983-41952022000500003
Pereira, F.C., Borysov, S.S., 2019. Machine learning fundamentals. In: Antoniou, C., Dimitriou, L., Pereira, F. (eds.), Pereira Big Data and Transport Analytics. Elsevier, Amsterdam, The Netherlands, Chapter 2, pp. 9–29, ISBN 978-0-12-812970-8.
Pone, J., Kamau, Ash, A., Kamau, J., Hyndman, F., 2018. Palm oil fuel ash as a cement replacement in concrete. Modern Approach Mater. Sci. 1 (1), 4–8. https://doi.org/10.32474/MAMS.2018.01.000102
Salami, B.A., Olayiwola, T., Oyehan, T.A., Raji, I.A., 2021. Data-driven model for ternary-blend concrete compressive strength prediction using machine learning approach. Constr. Build. Mater. 301, 124152. https://doi.org/10.1016/j.conbuildmat.2021.124152
Salehi, H., Burgueno, R., 2018. Emerging artificial intelligence methods in structural engineering. Eng. Struct. 171 (2018), 170–189. https://doi.org/10.1016/j.engstruct.2018.05.084
Saridemir, M., Topcu, I.B., Ozcan, F., Severcan, M.H., 2009. Prediction of long-term effects of GGBFS on compressive strength of concrete by artificial neural networks and fuzzy logic. Constr. Build. Mater. 23 (3), 1279–1286. https://doi.org/10.1016/j.conbuildmat.2008.07.021
Sas, Z., Doherty, R., Kovacs, T., Soutsos, M., Sha, W., Schroeyers, W., 2017. Radiological evaluation of by-products used in construction and alternative applications; Part I. Preparation of a natural radioactivity database. Constr. Build. Mater. 150, 227–237. https://doi.org/10.1016/j.conbuildmat.2017.05.167
Sas, Z., Sha, W., Soutsos, M., Doherty, R., Bondar, D., Gijbels, K., Schroeyers, W., 2019. Radiological characterization of alkali-activated construction materials containing red mud, fly ash and ground granulated blast-furnace slag. Sci. Total Environ. 659, 1496–1504. https://doi.org/10.1016/j.scitotenv.2019.01.006
Serin, S., Morova, N., Terzi, S., Sargin, S., 2011. Determining amount of bituminous effects on asphalt concrete strength with artificial intelligence and statistical analysis methods. In: Int. Symp. Innov. Intellig. Syst. Appl. pp. 329–34. ⟨https://doi.org/10.1109/INISTA.2011.5946139⟩.
Shamsabadi, E.A., Roshan, N., Hadigheh, S.A., Nehdi, M.L., Khodabakhshian, A., Ghalehnovi, M., 2022. Machine learning-based compressive strength modelling of concrete incorporating waste marble powder. Constr. Build. Mater. 324, 126592. https://doi.org/10.1016/j.conbuildmat.2022.126592
Shehab, M., Abualigah, L., Shambour, Q., Abu-Hashem, M.A., Shambour, M.K.Y., Alsalibi, A.I., Gandomi, A.H., 2022. Machine learning in medical applications: a review of state-of-the-art.
Siddika, A., Al-Mamun, Md.A., Ali, Md.H., 2018. Study on concrete with rice husk ash. Innov. Infrastruct. Sol. 3, 18. https://doi.org/10.1007/s41062-018-0127-6
Solak, S., Turhan, S., Uqur, F.A., Goren, E., Gezer, F., Yeqingil, Z., Yeqingil, I., 2014. Evaluation of potential exposure risks of natural radioactivity levels emitted from building materials used in Adana, Turkey. Indoor Built Environ. 23, 594–602. https://doi.org/10.1177/1420326X12448075
Song, H., Ahmad, A., Farooq, F., Ostrowski, K.A., Maslak, M., Czarnecki, S., Aslam, F., 2021. Predicting the compressive strength of concrete with fly ash admixture using machine learning algorithms. Constr. Build. Mater. 308, 125021. https://doi.org/10.1016/j.conbuildmat.2021.125021
Thilakarathna, P.S.M., Seo, S., Baduge, K.S.K., Lee, H., Mendis, P., Foliente, G., 2020. Embodied carbon analysis and benchmarking emissions of high and ultra-high strength concrete using machine learning algorithms. J. Clean. Prod. 262, 121281. https://doi.org/10.1016/j.jclepro.2020.121281
Trevisi, R., Leonardi, F., Risica, S., Nuccetelli, C., 2018. Updated database on natural radioactivity in building materials in Europe. J. Environ. Radioact. 187, 90–105. https://doi.org/10.1016/j.jenvrad.2018.01.024
UNSCEAR, 1993. Sources and Effects of Ionizing Radiation, United Nations Scientific Committee on the Effects of Atomic Radiation UNSCEAR 2000 Report to the General Assembly, with Scientific Annexes. UNSCEAR 1993 Report, New York, NY. Available at: ⟨https://www.unscear.org/docs/publications/1993/UNSCEAR_1993_Report.pdf⟩, (Accessed 4 September 2022).
UNSCEAR, 2000. Sources and Effects of Ionizing Radiation, United Nations Scientific Committee on the Effects of Atomic Radiation UNSCEAR 2000 Report to the General Assembly, with Scientific Annexes. UNSCEAR 2000 Report, New York, NY. Available at: ⟨https://doi.org/10.1097/00004032-199907000-00007⟩, (Accessed 2 September 2022).
UNSCEAR, 2008. Effects of Ionizing Radiation: Report to the General Assembly, with Scientific Annexes, vol 1. United Nations, New York. Available at: ⟨https://www.unscear.org/docs/publications/2008/UNSCEAR_2008_Report_Vol.I.pdf⟩, (Accessed 5 September 2022).
WHO, 2009. WHO Handbook on Indoor Radon. World Health Organization, Geneva, Switzerland. Available at: ⟨http://apps.who.int/iris/bitstream/handle/10665/44149/9789241547673_eng.pdf;jsessionid=A9F3F418B142132D746F2B8FA8069CB7?sequence=1⟩, (Accessed 16 September 2022).
Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim. Res. 30 (1), 79–82. ⟨http://www.jstor.org/stable/24869236⟩.
Yan, N., Chen, X., 2015. Sustainability: don't waste seafood waste. Nature 524, 155–157. https://doi.org/10.1038/524155a
Yang, S.S., Xu, J., Yao, G.Z., 2004. Concrete strength evaluation based on fuzzy neural networks. In: August Proceedings of 2004 Int. Conf. Mach. Learn. Cyber. vol. 6, pp. 3344–47. ⟨https://doi.org/10.1109/icmlc.2004.1380356⟩.
Yeh, I.C., 1998. Modeling of strength of high-performance concrete using artificial neural networks. Cem. Concr. Res. 28 (12), 1797–1808. https://doi.org/10.1016/s0008-8846(98)00165-3
Ziolkowski, P., Niedostatkiewicz, m., 2019. Machine learning techniques in concrete mix design. Materials 12 (8), 1256. https://doi.org/10.3390/ma12081256