PRELIMINARY PREDICTION OF RADIONUCLIDE INVENTORY IN NUCLEAR WASTE USING MACHINE LEARNING

Do-Hyeon Kim and Jun-Yeop Lee*

School of Mechanical Engineering, Pusan National University

2, Busandeahak-ro 63beon-gil, Geumjeong-gu, Busan 46241, Republic of Korea

** E-mail: jylee@pusan.ac.kr*

ABSTRACT

This study develops a model to predict the concentration of difficult-to-measure (DTM) radionuclides using a machine learning-based XGBoost (XGB) method, leveraging the Electric Power Research Institute (EPRI) database. A dataset comprising four input variables, considered easy-to-measure (ETM) radionuclides, including plant number (PLT) and the concentrations of ^{60}Co, ^{137}Cs, and ^{144}Ce, was constructed for use in the XGB prediction model. The model was trained using nested cross-validation, and its performance was assessed through the coefficient of determination (R^2) and root mean square error (RMSE). The results demonstrate that the model provides reasonable predictive accuracy for both metal and actinide radionuclides. This machine learning approach offers a novel solution to challenges in nuclear waste management and can also be used to cross-validate the concentrations of DTM radionuclides obtained through other methods.

KEY WORDS

Scaling factor, XGBoost, ETM, DTM

1. INTRODUCTION

For the final disposal of radioactive waste from nuclear power plants (NPPs), it is crucial to have accurate information about the characteristics and quantities of radionuclides contained in the radioactive waste packages. Most radionuclides that are significant for the long-term storage of low- and intermediate-level waste are difficult-to-measure (DTM) radionuclides, such as ^{3}H, ^{129}I, ^{63}Ni, ^{90}Sr, ^{238}Pu, etc.

Since many DTM radionuclides cannot be easily or directly measured, the scaling factor (SF) method has been widely employed for the safe management of nuclear waste. The SF method is a technique used to estimate the concentrations of DTM radionuclides based on the concentrations of easy-to-measure (ETM) radionuclides, assuming a correlation between the ETM and DTM radionuclides. The formula for SF is as follows:

$$SF = A^{DTM}/A^{ETM} \quad (1)$$

where A^{DTM} is the activity of DTM radionuclide, and A^{ETM} is the activity of ETM radionuclide. Note that the derivation of scaling factors (SFs) has relied on empirical models in general, which are often limited to simple parametric methods.

The objective of this study is to develop a model for predicting the concentrations of DTM radionuclides based on the concentrations of ETM radionuclides. The prediction model was developed using the machine learning-based XGBoost (XGB) method[1], coupled with a nested cross-validation approach.

2. MATERIALS AND METHODS

2.1 Data collection and pre-processing

Radionuclide inventory of nuclear waste data from EPRI-5077[2], which includes the quantitative information of both DTM and ETM radionuclides, was employed to construct the prediction model. Plant number was included as input data to distinguish the source of the radioactive waste from each nuclear power plant. Three ETM radionuclides (^{60}Co, ^{137}Cs, and ^{144}Ce) were used as input data, while the output data consisted of 13 DTM radionuclides (^{3}H, ^{14}C, ^{55}Fe, ^{63}Ni, ^{90}Sr, ^{99}Tc, ^{129}I, $^{238+239+241}$Pu, ^{241}Am, and $^{242+244}$Cm).

Since the radionuclide concentrations were represented at very low levels, making them difficult to differentiate on a linear scale, a logarithmic scale was applied to enhance the model performance.

2.2 Machine learning model

For the development of the machine learning model, the XGBoost (XGB) method was adopted in this study. XGB is a supervised ensemble machine learning approach based on multiple decision trees with various hyperparameters and is well-known for its high performance in regression tasks and its ability to handle imbalanced data.

To address overfitting and data bias issues during the data-splitting process, nested cross-validation (NCV)[3] was employed for model tuning. In this study, the NCV method utilized a double-loop structure, with five folds in both the inner and outer loops (referred to as 5-5 NCV), to enhance robustness and provide redundancy of the model.

Additionally, Bayesian optimization[4] was applied for hyperparameter tuning to efficiently identify the optimal set of hyperparameters. This technique is a probabilistic model-based optimization method that leverages a surrogate model to predict and explore the hyperparameter space more effectively.

Finally, the performance of the prediction model developed in this study was evaluated using R^2 and RMSE metrics.

3. RESULTS AND DISCUSSION

Table 1 lists performance results obtained from the 5-5 NCV on the DTM radionuclides, representing the average R^2 and RMSE values from the five folds of the outer loop.

Table 1: Performance evaluation results for the selective DTM radionuclides derived in this study.

Volatile or mobile radionuclide group				
	^{3}H	^{14}C	^{99}Tc	^{129}I
R^2	0.358	0.578	0.516	0.395
RMSE	1.017	1.095	0.857	0.895
Metal or immobile radionuclide group*				
	^{55}Fe	^{63}Ni	^{238}Pu	^{241}Am	^{244}Cm
R^2	0.851	0.887	0.772	0.814	0.805
RMSE	0.861	0.732	0.830	0.759	0.804

*Only a selected subset of radionuclides is presented.

According to the results, the DTM radionuclides were categorized into two groups. The volatile or mobile group (referred to as Group 1) included radionuclides such as ^{3}H, ^{14}C, ^{99}Tc, and ^{129}I, which exhibited relatively lower performance, with R^2 values ranging from 0.358 to 0.578 and RMSE values from 0.857 to 1.095. In contrast, the metal or immobile group (referred to as Group 2) comprised ^{55}Fe, ^{63}Ni, ^{90}Sr, $^{238+239+241}Pu$, ^{241}Am, and $^{242+244}Cm$, demonstrating higher performance, with R^2 values ranging from 0.772 to 0.887 and RMSE values from 0.732 to 0.862.

Fig. 1 compares the experimental concentrations with the predicted concentrations derived by integrating the five folds of the outer loop. The prediction for ^{3}H exhibits relatively low accuracy, with an R^2 value of 0.381, whereas ^{63}Ni shows strong consistency, achieving an overall R^2 value of 0.887. These results confirm that integrating the predicted values into a single output produces outcomes that are relatively consistent with those presented in Table 1.

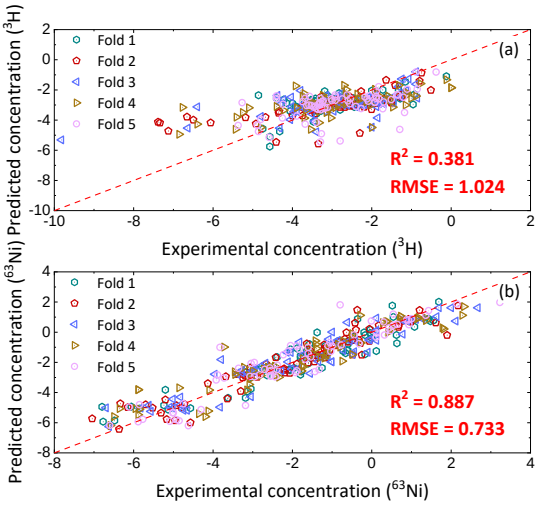

Fig. l: Comparison of prediction performance between (a) ^{3}H and (b) ^{63}Ni. The dashed line indicates the ideal fit where the predicted values are equal to observed values.

As shown in Table 1, radionuclides in Group 1 exhibit either volatile[5] or anionic characteristics, which are anticipated to pose challenges in predicting their concentrations. In contrast, Group 2 radionuclides, consisting of metals and actinides, are characterized by their low solubility[6], restricting their mobility and presumably contributing to their relatively stable presence within the waste matrix. According to the result obtained in this study, these properties are expected to enhance the accuracy of prediction performance.

4. CONCLUSION

A computational model for reliably predicting the concentrations of DTM radionuclides based on the concentrations of ETM radionuclides was developed using the machine learning-based XGBoost (XGB) model. The R^2 values for radionuclides in Group 1 were below 0.6, whereas those in Group 2 exceeded 0.75, demonstrating better performance.

The model established in this study enables the evaluation of DTM radionuclide quantities in radioactive waste under arbitrary concentrations of ETM radionuclides. Furthermore, this model is expected to overcome the limitations of the classical scaling factor method through a novel approach to addressing challenges in nuclear waste management. It can also be used to cross-validate the concentrations of DTM radionuclides determined through other methods.

ACKNOWLEDGEMENTS

This work was supported by the Institute for Korea Spent Nuclear Fuel (iKSNF) and National Research Foundation of Korea (NRF) grant funded by the Korea government [No. NRF-2021M2E1A1085204] and was supported by the 'Human Resources Program in Energy Technology' of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), which was funded by the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea [No. RS-2024-00398425].

REFERENCES

1 T. Chen and C. Guestrin, *Proc. 22nd acm sigkdd international conference on knowledge discovery and data mining* (2016) p.785–794.
2 W.T. Best, Electric Power Research Institute Report, EPRI-5077 (1987)
3 G. C. Cawley and N.L. Talbot, *J. Mach. Learn. Res.* 11 (2010) 2079–2107.
4 B. Shahriari et al., *Proc. IEEE 104*, (2016) p.148–175.
5 IAEA, Nuclear Energy Series No. NW-T-1.7 (2019)
6 IAEA, Nuclear Energy Series No. NW-T-1.18 (2009)