Machine Learning-Based Exploration of High-Entropy LDH for Iodate Adsorption

2025 Regeneron ISEF
Jiyeong Jeong, Bokyoung Joo, One Choi
Korea Science Academy of KAIST

MARCH 24, 2025

Abstract

Nuclear fission energy accounts for 10% of the total energy generated. As it does not emit greenhouse gases, it is considered a leading sustainable energy source. However, radioactive anions can be released through sewage, causing mutations and cancer in living organisms, which leads to severe disruption in marine ecosystems. Specifically, radioactive iodine is one of the seven long-lived nuclear fission products, with a half-life of 15.7 million years. A basic Layered Double Hydroxide (LDH) was previously developed to adsorb radioactive iodine. To address its limitations, the high-entropy effect was proposed as a solution to the challenges associated with using the basic LDH structure. High-entropy materials are composed of various elements, allowing the design of materials with high efficiency and stability suitable for specific applications. However, this approach results in an immense number of possible compositions to explore, and the limited experimental data reported so far restricts conventional exploration methods. In this study, machine learning was utilized to explore quaternary and quinary LDH compositions for radioactive iodine adsorption. As a result, Cu1(CrFeAl)1 was identified as the LDH composition with the highest adsorption performance, increasing the previous maximum performance of 89.1% to 91.8%, all while exploring less than 1% of the total experimental space. This work demonstrates and analyzes high-entropy LDH with outstanding iodate adsorption performance and validates the efficiency and feasibility of this approach. It has the potential to be a pioneer in anionic pollutant adsorption materials, ultimately contributing to environmental preservation.

1 Introduction

1.1 Research Background and Purpose

In the nuclear fission process of ^{235}U, radioactive isotopes of iodine, ^{131}I and ^{129}I, are generated as byproducts. The record ^{129}I is one of the seven long-lived fission products. If radioactive iodine is released to the outside, the high energy of the radiation can have a serious impact on the food chain of the ecosystem, greatly increasing the possibility of genetic mutations. The same applies if it enters the body; the possibility of causing cancer, such as thyroid cancer, increases significantly. ^{129}I has extremely long half-life of approximately 15.7 million years. Thus, once it is released, natural decomposition hardly occurs without an artificial adsorption process.

Radioactive iodine exists in the form of iodide ion (I^-) or iodate (IO_3^-) in radioactive wastewater. Unlike I^-, no substance can effectively remove IO_3^-. Attempts were made to use silver-based zeolites or activated carbon to adsorb it, but there are clear limitations in terms of cost efficiency, environmental safety, and adsorption efficiency [1]. To solve this, layered double hydroxide (LDH) was selected as the adsorption material in this study, and high entropy ceramic (HEC) made of various elements was utilized to further increase the adsorption capacity. Lastly, machine learning (ML) was used to secure the reality and efficiency of the research while also conducting the research. This study presents a new methodology for developing an adsorbent for radioactive IO_3^- by utilizing high-entropy materials and machine learning. This is expected to be an effective approach to overcome the limitations of existing adsorption technologies and solve the problem of environmental pollution caused by radioactive substances.

1.2 Theoretical Backgrounds

1.2.1 Layered Double Hydroxide (LDH)

Layered double hydroxide (LDH) is a material with a unique structure in which cation layers and anion layers are alternately stacked as shown in Figure 1. Having excellent anion exchange capacity, it exhibits high adsorption performance. LDH forms an octahedral structure, where a central metal ion in the cation layer is surrounded by OH^- and can incorporate over 15 metal elements [2].

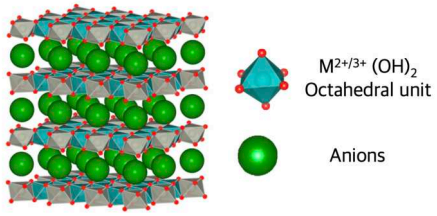

Figure 1: Layered Double Hydroxide

$$\left[M^{2+}_{1-x}M^{3+}_{x}(OH)_2\right]^{x+}(A^{n-})_{x/n}\cdot mH_2O \quad (1)$$

Here, M^{2+} and M^{3+} are divalent and trivalent metal cations, respectively, and A is an interlayer anion. These structural characteristics of LDH suggest the possibility that it is suitable for adsorbing anionic substances such as radioactive iodate (IO_3^-). However, previous adsorption studies using

LDH have pointed out the limitations of the insufficient adsorption capacity and the reversal of selectivity, where other anions are preferentially adsorbed over iodate.

1.2.2 High Entropy Ceramics (HECs)

High entropy ceramics refer to systems in which five or more elements are each composed of 5-35 atomic%. In terms of compositional entropy, it refers to materials with an entropy value of 1.6R or higher, where R is the gas constant.

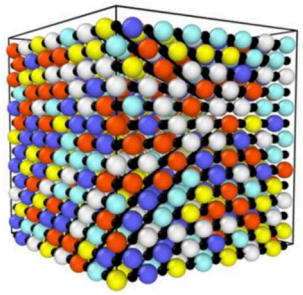

Figure 2: High Entropy Ceramics

These high-entropy ceramics can exhibit superior properties compared to other materials based on high entropy. High thermal and structural stability and being multifunctional are such properties. These properties originate from high-entropy effect, cocktail effect, grid distortion effect, and sluggish diffusion effect. Thus, it is expected to highly enhance the performance when integrated with LDH [3].

1.2.3 Machine Learning (ML)

Machine learning (ML) is a technology that identifies and analyzes correlations in massive data to predict results. It learns testing data, creates multidimensional, multivariable functions that represent tendencies, and performs predictions by fitting the functions to inputs. The two machine learning methods used in this study are Classification and Regression [4].

The classification model dichotomizes data according to criteria and defines a decision boundary while learning training data. It produces an output, which is a classification result for inputs that have not yet been determined. The classification model mainly evaluates the validity through indicators such as accuracy, precision, recall, and F1 score. Representative classification models are Decision Tree, Random Forest, and Support Vector Machine, and these three models were utilized in this study.

A decision tree divides data into two based on the question that can best distinguish the data in a single classification, and determines the area to which the variable corresponds by repeating this in each category. Random forest classification is a method that uses multiple decision trees to generate and train a certain number of classifiers that use the same algorithm to make decisions [5]. A support vector machine (SVM) classifies data by determining the support vector that can best divide input data with different values.

The regression model, on the other hand, learns the relationship between x and y of the training data, generating a multidimensional vector function with real values as output values. It is used to predict continuous data. The validity evaluation measures of the regression model include the mean absolute error, RMSE, and coefficient of determination (R2). Representative regression models are Random Forest [Equation (2)], Linear [Equation (3)], Ridge [Equation (4)], and Support Vector Regression [Equation (5)], and these four regression models were utilized in this study.

$$\hat{y}_{\mathrm{RF}} = \frac{1}{N}\sum_{i=1}^{N} Ti(x) \tag{2}$$

$$\min_{\mathbf{w},b} \sum_{i=1}^{n} (y_i - \mathbf{w}^\top \mathbf{x}_i - b)^2 \tag{3}$$

$$\min_{\mathbf{w},b} \sum_{i=1}^{n} (y_i - \mathbf{w}^\top \mathbf{x}_i - b)^2 + \lambda \|\mathbf{w}\|^2 \tag{4}$$

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 \quad \text{s.t.} \quad y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \quad \forall i \tag{5}$$

Machine learning is very effective in analyzing complex and massive data that humans cannot handle, and predicting new data or making decisions based on this. In this study, we utilized the advantages of machine learning to predict adsorption capacity in complex multi-element systems and explore optimal combinations. Through machine learning, we were able to efficiently explore the astronomical combination possibilities that arise from the various compositions of the central metals in the LDH cation layer.

2 Research Methods

2.1 Synthetic Composition Selection Using Machine Learning (ML)

ML model code was written to predict quaternary and quinary compositions with high IO_3^- adsorption performance. The models were trained with binary and ternary LDH data of synthesizability index (SI) and adsorption capacity. These model then predict the synthesizability, and among the quaternary and quinary composition that are expected to be synthesized using classification models, IO_3^- adsorption capacity of them was predicted using regression models. Refer to Figure 3 as an example. Previously mentioned models were used and the validity of them was precisely analyzed using the indicators.

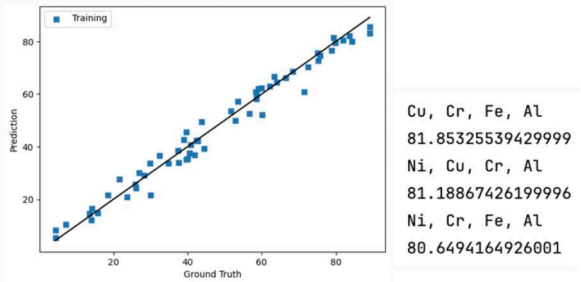

Figure 3: Output example of machine learning results

Among a total of 12 model combinations using three classifications and four regression models, the quaternary and quinary compositions showing the highest adsorption performance for each model were selected. [Table 1-2].

Table 1. Composition with the highest predicted adsorption capacity and its value in quaternary system

Classification Model	Regression Model			
	Random Forest	Ridge	Linear	SVM
Random Forest	Cu, Cr, Fe, Al	Ni, Cu, Cr, Al	Ni, Cu, Cr, Al	Ni, Cu, Cr, Al
	81.85	93.67	98.90	86.48
Decision Tree	Ni, Cu, Cr, Al	Ni, Cu, Cr, Al	Ni, Cu, Cr, Al	Ni, Cu, Cr, Al
	86.48	93.67	98.90	86.48
SVC	Cu, Cr, Fe, Al	Ni, Cu, Cr, Al	Ni, Cu, Cr, Al	Ni, Cu, Cr, Al
	81.97	93.67	98.90	86.48

Table 2. Composition with the highest predicted adsorption capacity and its value in quinary system

Classification Model	Regression Model			
	Random Forest	Ridge	Linear	SVM
Random Forest	Ni, Cu, Cr, Fe, Al	Co, Ni, Cu, Cr, Al	Co, Ni, Cu, Cr, Al	Co, Ni, Cu, Cr, Al
	80.45	91.62	97.47	90.97
Decision Tree	Ni, Cu, Cr, Fe, Al	Co, Ni, Cu, Cr, Al	Ni, Zn, Cr, Al, Bi	Co, Ni, Cu, Cr, Al
	80.85	91.62	102.25	90.97
SVC	Ni, Cu, Cr, Fe, Al	Co, Ni, Cu, Cr, Al	Co, Ni, Cu, Cr, Al	Co, Ni, Cu, Cr, Al
	80.85	91.62	97.47	90.97

As a result, the quaternary composition (Ni, Cu, Cr, Al) and (Cu, Cr, Fe, Al) were predicted to have the highest adsorption performance, along with the quinary composition (Co, Ni, Cu, Cr, Al), (Ni, Cu, Cr, Fe, Al), and (Ni, Zn, Cr, Al, Bi). Here, the composition (Ni, Zn, Cr, Al, Bi) was predicted not to be synthesized empirically. Therefore, these four were selected as compositions to be verified through actual experiments. For a control group, (Ni, Cu, Fe, Al) and (Ni, Zn, Cr, Fe, Al), selected through a trial-and-error method, were used.

Table 3. Selected compositions

No.	Composition	Method
#4.1	(Ni, Cu, Cr, Al)	ML
#4.2	(Cu, Cr, Fe, Al)	ML
#4.3	(Ni, Cu, Fe, Al)	Trial-and-Error
#5.1	(Co, Ni, Cu, Cr, Al)	ML
#5.2	(Ni, Cu, Cr, Fe, Al)	ML
#5.3	(Ni, Zn, Cr, Fe, Al)	Trial-and-Error

2.2 Ceramic synthesis of six compositions

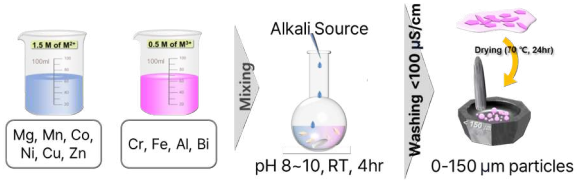

Figure 4: Schematic diagram of co-precipitation method

1.5 M divalent metal hydrate (Co, Ni, Cu, Zn) aqueous solution, 0.5 M trivalent metal hydrate (Cr, Fe, Al) aqueous solution, 1 M NaOH, and 0.5 M Na_2CO_3 basic solutions to be used in the synthesis were prepared.

LDH ceramic synthesis was performed using the coprecipitation method–most common LDH synthesis method. Schematic diagram of it is expressed in Figure 4. At this stage, the ratio of $[M^{2+}]$ to $[M^{3+}]$ is 3 : 1, and metals of the same valence must be equimolar. Maintaining a 3 : 1 concentration ratio, 10 mL of each solution was combined, totaling 20 mL.

After thoroughly stirring the prepared solution with a magnetic bar, the pH was adjusted to 8–10, the optimal range for LDH synthesis, using a pre-prepared alkaline solution (1 M NaOH, 0.5 M Na_2CO_3) [10, 11]. The solution was then stirred for 4 hours. Each of the three pots synthesized for each composition was combined, and the solution was washed 7–8 times until the ion conductivity of the supernatant was reduced to below 100 μS/cm. The samples were then dried in a vacuum oven at 70°C for 24 hours. The dried samples were ground using a mortar and pestle to standardize the particle size to less than 150 μm, then collected and stored in a vial. Refer to Figure 4 for representation of the process.

Composition	Alkali Volume	$pH_{initial}$	pH_{final}
$(NiCu)_3(CrAl)_1$	23 mL	8.94 ± 0.17	7.80 ± 0.24
$Cu_3(CrFeAl)_1$	23 mL	8.61 ± 0.28	8.26 ± 0.23
$(NiCu)_3(FeAl)_1$	23 mL	8.37 ± 0.39	7.69 ± 0.24
$(CoNiCu)_3(CrAl)_1$	21.5 mL	8.76 ± 0.09	7.25 ± 0.24
$(NiCu)_3(CrFeAl)_1$	21 mL	7.75 ± 0.55	7.19 ± 0.29
$(NiZn)_3(CrFeAl)_1$	21.5 mL	8.58 ± 0.31	6.82 ± 0.56

Table 4. Chemical Composition Table (Category as Columns)

2.3 Verification of LDH synthesis through multiple analyses

XRD analysis was performed to verifty that composite was synthesized as the target LDH phase [12]. Measurements were taken once every 0.01° at a rate of 1° per 10.0s from 5° to 80°. In order to avoid charge accumulation, approximately 13 nm of platinum coating (5 mA, 200 s $\rightarrow$ 13 nm) was performed and then analyzed [15]. SEM-EDS analysis was performed to evaluate the surface shape of the composite and whether each metal component was evenly distributed within the powder and synthesized [13-14]. ICP-OES analysis was performed to verify the mole ratio of $[M^{2+}]$: $[M^{3+}]$

is 3 : 1 and if the metals of the same valence were equimolar [16].

2.4 IO_3^- adsorption experiment

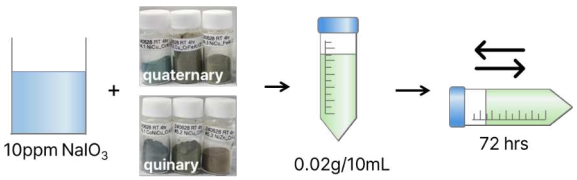

Figure 5: Schematic Diagram of Iodate Adsorption Process

First, a 10 ppm IO_3^- solution was prepared. The concentration of the prepared NaIO3 solution was re-verified as 10 ppm using UV-Vis spectroscopy. In addition, 0.02 g (±0.001 g) of the synthesized sample was placed in four 15 mL tubes (1 DI water + 3 10 ppm IO_3^- solutions) for each composition. 10 mL of 10 ppm NaIO3 solution was placed in each 15 mL tube, and adsorption was performed in a stirrer for 72 hours. After the adsorption test, the treated water was filtered through a 0.45 um PTFE filter for each sample to separate it from the remaining powder and stored in a new conical tube.

2.5 Evaluation of the IO_3^- adsorption performance of synthesized quaternary and quinary LDHs through various analyses

SEM and EDS mapping were performed to confirm how IO_3^- was spatially adsorbed in the LDH sample. To confirm whether new bonds were formed by IO_3^- adsorption, the binding environment was analyzed through Raman spectroscopy [17]. In order to quantitatively evaluate the adsorption performance of IO_3^- of each LDH composition, the residual concentration of IO_3^- was measured through ion chromatography.

2.6 Finding the most accurate ML model and improve ML models

The most accurate ML model was explored by examining the RMSE and the error between the predicted and experimental values for both classification and regression models. To address imperfect accuracy Bayesian hyperparameter was incorporated. Additionally, the SI and adsorption capacity of the synthesized quaternary/quinary compositions was added to the input data to evaluate the improvement in the model's predictions. Furthermore, by incrementally adding ternary data points to the binary training data, the evolution of prediction accuracy for the ternary data was examined, thereby investigating whether the dimensional gap could be overcame.

3 Research Results

3.1 Results for LDH crystal phase

The crystal structure of the sample was confirmed using X-Ray Diffraction (XRD). The results of XRD analysis on the six synthesized LDHs are shown in Figure 6. The XRD intensity of each sample was normalized to a value between 0 and 1 to be compared directly.

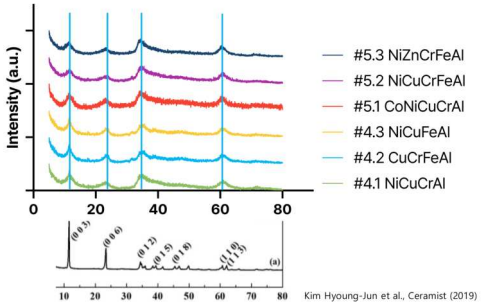

Figure 6: Comparison of XRD results and peaks

Figure 6 shows the XRD analysis results for six samples. Each peak is a unique characteristic depending on the crystal phase of the material. As a result of our analysis, each synthesized quaternary and quinary LDH was found to match the LDH phase in the database ($Cu_6Al_2(OH)_{16}CO_3 \cdot 4\,H_2O$, PDF card #: 00-037-0630). However, it should be noted that in this experiment,

#4.2 $(Cu)_3(CrFeAl)_1$ was manufactured in the malachite phase and the analysis was conducted using a $(Cu)_3(CrFeAl)_1$ LDH sample previously manufactured under the same conditions. Here, it was confirmed that the LDH compounds was successfully synthesized as all representative peaks were identical to the peaks of LDH in the database, and adsorption experiments were performed thereafter.

3.2 Evaluation of surface shape and element loading of LDH

The surface shape of the material was confirmed using scanning electron microscopy (SEM), and the distribution and concentration of each element were evaluated using energy dispersive X-ray spectroscopy (EDS). If a specific element was charged locally, it cannot be said to have been synthesized properly, and this may affect the adsorption capacity. (Figure 7-8) From Figure 7-8, it can be confirmed that the elements are evenly distributed, and the at.% value of iodine (I) has increased after adsorption compared to before adsorption. In other words, it can be seen that the synthesized materials adsorb IO_3^-.

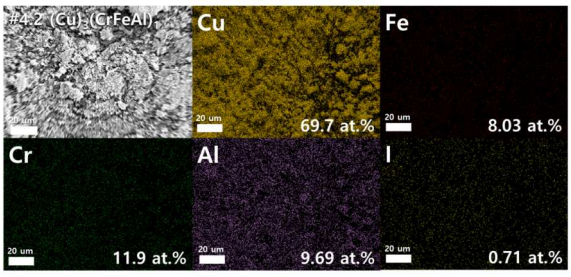

Figure 7: SEM-EDS Result of $Cu_3(CrFeAl)_1$ before adsorption

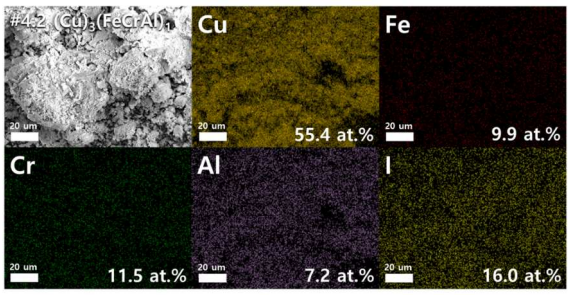

Figure 8: SEM-EDS Result of $Cu_3(CrFeAl)_1$ after adsorption

3.3 Evaluation of binding within LDH

To prepare a solution for adsorption on ceramics, a $NaIO_3$ solution was made, and its concentration was verified using UV-Vis spectroscopy. Standard solutions of $NaIO_3$ (3 ppm, 6 ppm, 12 ppm) were used to create a calibration curve. After confirming the accuracy of the 10 ppm $NaIO_3$ solution, the adsorption experiment was conducted.

3.4 Quantitative analysis of metal elements in LDH

Inductively coupled plasma spectroscopy (ICP-OES) was utilized to evaluate whether the ratio of each metal in the synthesized LDH was as designed. ICP-OES uses high-temperature argon plasma of over 10,000 K to excite electrons and calculates the wavelength of electromagnetic waves emitted. This allows it to classify and analyze elements accurately.

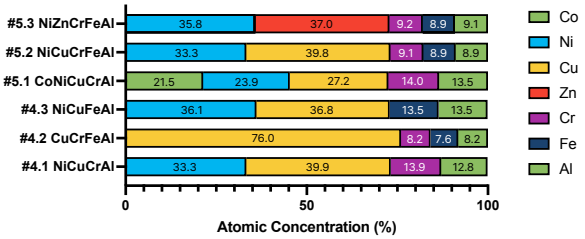

Figure 9: Ratio of elements in compositions

The ratio of elements by composition obtained through ICP-OES and they are represented in Figure 9. This shows that most divalent metals reach around 75 at.% in all samples. For more specific analysis, we quantitatively examined how much the experimentally obtained ratio between metals differed from the theoretical ratio.

Sample	Composition	Atomic Percent (%)		
		M^{2+}	M^{3+}	Total
#4.1	$(NiCu)_3(CrAl)_1$	73.2	26.8	100
#4.2	$Cu_3(CrFeAl)_1$	76.0	24.0	100
#4.3	$(NiCu)_3(FeAl)_1$	72.9	27.1	100
#5.1	$(CoNiCu)_3(CrAl)_1$	72.5	27.5	100
#5.2	$(NiCu)_3(CrFeAl)_1$	73.1	26.9	100
#5.3	$(NiZn)_3(CrFeAl)_1$	72.8	27.2	100

Table 5. Atomic percent of M^{2+} and M^{3+} in each composition

When comparing the amount of metal elements actually added and the amount of metal elements obtained experimentally for each composition, an error of $\pm 2.59\%$ occurred on average in the case of divalent metals, and an error of $\pm 7.39\%$ occurred in the case of trivalent metals. It can be considered as a insignificant experimental error.

Although it was shown that the divalent and trivalent metals were appropriately distributed in a ratio of 3 : 1, it is also necessary to evaluate whether they were charged in equimolar amounts within each coordination number. Table ?? shows the percent difference between the actual mixed ratio and the experimental ratio for each element. Mostly, it shows a stable error range of less than $\pm 10\%$, a reasonable, insignificant error that could occur experimentally.

3.5 Comparison of binding environments before and after adsorption in synthesized LDH

Raman spectroscopy can determine whether IO_3^- is adsorbed. It also determines the change in the binding environment before and after adsorption by a generated signal (peak position) according to Raman shift (x-axis). According to Figure 11, a new peak has appeared at wavelength from 770 nm to 850 nm. This indicates that IO_3^- has binded to the material. To represent the result, the value of the detection signal was normalized to a value between 0 and 1 for relative comparison of each sample.

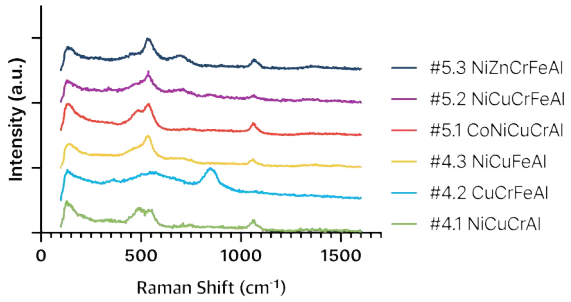

Figure 10: Raman Spectroscopy of LDH Before Adsorption

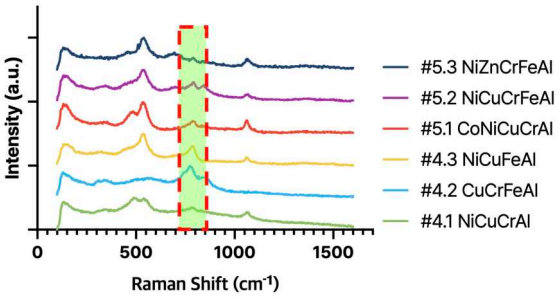

Figure 11: Raman Spectroscopy of LDH Before Adsorption

3.6 Results for LDH crystalline phase after adsorption

The XRD results after adsorbing IO_3^- are as follows, and it can be seen that even if IO_3^- is adsorbed within the LDH structure, there is no change in the crystal phase as you can notice in Figure 12.

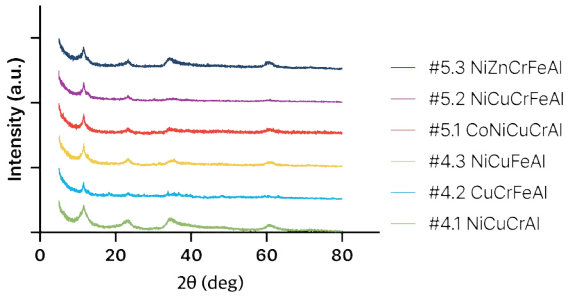

Figure 12: XRD of LDH After Adsorption

3.7 Quantitative adsorption performance of synthesized LDH

If it was confirmed that IO_3^- was properly adsorbed to LDH through the previous method, ion chromatography (IC) quantified the concentration of IO_3^- remaining in the solution and the adsorption capacity could be evaluated [18]. The evaluated adsorption capacity was organized as [Table 2].

No.	Composition	Adsorption Capacity
#4.1	$(NiCu)_3(CrAl)_1$	56.8±3.28 %
#4.2	$(CuCr)_3(FeAl)_1$	91.8±0.23 %
#4.3	$(NiCu)_3(FeAl)_1$	87.4±0.33 %
#5.1	$(CoNiCu)_3(CrAl)_1$	54.1±1.00 %
#5.2	$(NiCu)_3(CrFeAl)_1$	69.5±0.29 %
#5.3	$(NiZn)_3(CrFeAl)_1$	38.2±1.47 %

Table 6. Composition and adsorption capacity

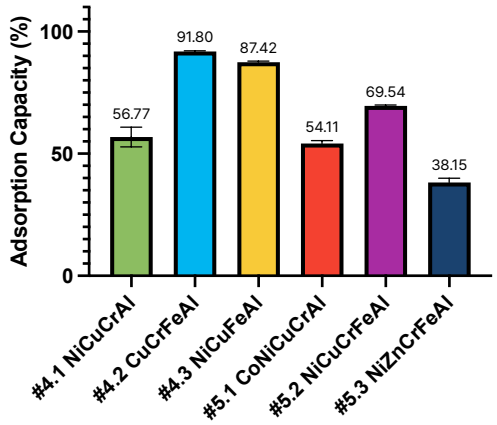

Figure 13: Adsorption Capacity of LDH

According to Figure 13, #4.2 $Cu_3(CrFeAl)_1$ showed the best adsorption capacity in the quaternary system, and #5.2 $(NiCu)_3(CrFeAl)_1$ in the quinary system. That is, it was found that the sample obtained through ML showed better adsorption performance compared to #4.3 $(NiCu)_3(FeAl)_1$ and #5.3 $(NiZn)_3(CrFeAl)_1$, which were intuitively selected based on binary and ternary data.

3.8 Selecting a machine learning model with the best prediction performance

The model with the highest accuracy was identified by comparing RMSE and relative error between the actual adsorption performance (experimental value) and the predicted performance (predicted value) for each model. At this stage, the errors for each relative composition in each model were averaged and utilized.

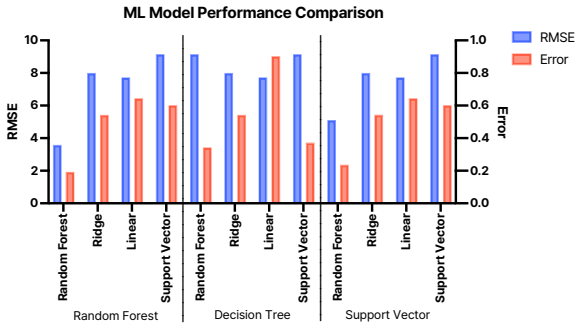

Figure 14: ML Model Performance Comparison

According to Figure 14, the results show that using Random Forest in both classification and regression has the smallest relative error and RMSE. In other words, it can be concluded that the Random Forest–Random Forest (RF-RF) model was the most accurate in predicting.

3.9 ML model improvement

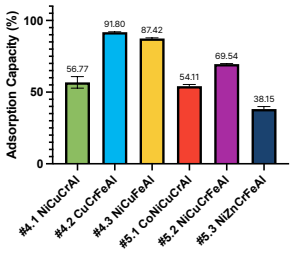

Figure 15: Adsorption Capacity of LDH

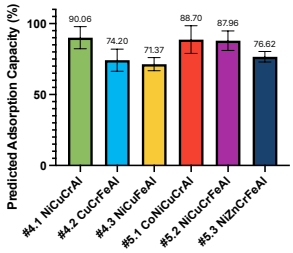

Figure 16: Predicted Adsorption Capacity

When comparing the IO_3^- adsorption capacity of the six samples predicted by the ML model with the experimentally measured values, a significant discrepancy was observed in Figure 16, indicating low accuracy of the ML model. This decrease in accuracy is likely due to the dimension mismatch between the training data (binary and ternary) and the output (quaternary and quinary). To address this, two approaches were implemented: adding data of the same dimension as the test data to the training set and utilizing an advanced model designed to enhance ML performance.

3.9.1 Addition of 6 Synthesized Samples

While the aim was to add over 20 SI and adsorption capacity data for quaternary and quinary systems, but no ad-

ditional IO_3^- adsorption capacity data was found after an extensive reviewing relevant papers. Consequently, we incorporated the IO_3^- adsorption capacity of the six synthesized samples into the training data and re-predicted their adsorption capacity.

3.9.2 Applying Existing Models for Enhanced Machine Learning Performance

To additionally enhance the ML model's accuracy, Bayesian hyperparameter was applied, which was known for improving ML performance.

3.10 Analysis of results of each method

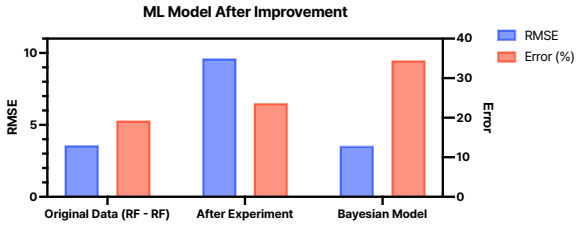

Figure 17: ML Model Performance Comparison

The RMSE and relative error rates values between the actual experimental results and predicted values of six samples were compared across three models: the RF-RF model before adding the six synthesized samples (Original Data), the RF-RF model after adding them (After Experiment), and the Bayesian hyperparameter model. The graph is expressed as in Figure 17.

Compared to the existing model and the RF-RF model, the Bayesian hyperparameter model showed no significant improvement in RMSE values, and the error rate between experimental and predicted values increased. This suggests that Bayesian hyperparameter tuning is not suitable for composition search.

In contrast, the After Experiment model (which incorporated the six tested compositions) showed an increased error rate and an RMSE value approximately three times higher than the RF-RF model, indicating a notable decline in accuracy. This is likely due to the limited data available for the quaternary and quinary systems (only six samples), as well as the possibility of outlier data that contradicts the overall trend within these compositions.

3.11 Possibility of improving the ML model by adding quaternary and quinary data

The performance of ML was insufficient due to the dimensional difference between the data sets. Consequently, alternative potential to improve the model as more data were added to predict were investigated. Since there was a limitation in quaternary and quinary data, this was verified by adding ternary data gradually to the training data set, which was originally consisted of only binary data, and observed the change in RMSE of the prediction. It confirms that improvement in prediction performance as the difference in dimension mitigates.

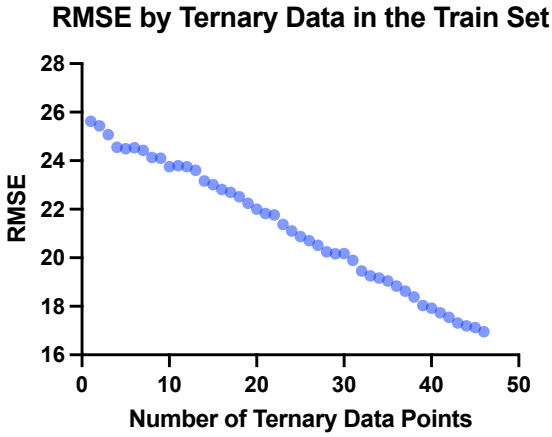

Figure 18: ML Model Performance Improvement

As a result, the RMSE of the model continued to decrease as ternary data was gradually added to the training data set. In particular, it was observed that the RMSE decreased sharply in the section until 20% of the ternary data was added. Through this, it can be seen that the accuracy of the model can be sufficiently improved by adding quaternary and quinary data to the training data of the current model, and that the model can be effectively improved by adding only 20% of the number of possible cases.

4 Conclusion

4.1 Research Conclusion

When looking at the compositions determined through machine learning and the ion chromatography results of the compositions determined intuitively by looking at the binary and ternary data, the IO_3^- adsorption performance of the compositions determined through ML tended to be higher. In addition, among the 440 compositions of the quaternary and quinary materials, only 4 compositions (less than 1%) were selected, but the composition $Cu_3(CrFeAl)_1$ with an adsorption performance of 91.8% was found, which is higher than the maximum adsorption performance of 89.2% of all the binary and ternary compositions. Through this, it can be seen that the method of selecting the composition using machine learning is very efficient and effective in terms of time and cost. It is expected that this material exploration method using ML will be useful in exploring the suitable composition of new materials that can be used for various pollutant adsorption, rockets, turbine engines, catalysts, etc. Moreover, this exploration method utilizing machine learning and high-entropy LDH is expected to be effectively applied when attempting a method to remove pollutants in the form of anions other than IO_3^-.

4.2 Limitations and Future Research

However, there was a difference between the IO_3^- adsorption capacity predicted by ML and the IO_3^- adsorption capacity measured through experiments, which showed that the accuracy of the ML model designed in this study was not high. This seems to be due to the difference in the dimensions of the training data and the test data in the ML model. However, the RMSE of the model predicting the ternary system continuously decreased as the ternary system data was gradually added to the training data set consisting of only the binary system. This showed that the accuracy of the model can be sufficiently improved by adding additional quaternary and quinary data to the training data of the current model. Additionally, it seems that research could be presented to identify the performance of compositions with low predicted adsorption capacity rather than compositions with high predicted adsorption capacity in the machine learning process, or to verify the IO_3^- selectivity of synthesized metal compositions.

References

[1] Srivastava, V., Dubey, A., & Shrivastava, R. (2021). Green Synthesis of Nanoparticles: A Step Towards Sustainability. *SN Applied Sciences*, 3(1), 1–13.

[2] Oses, C., Curtarolo, S., & Levy, O. (2020). High-Entropy Ceramics. *Nature Reviews Materials*, 5, 295–309.

[3] Saju, D., Mohanan, P., & Thomas, S. (2020). Layered Double Hydroxides: Fundamentals to Applications. In *Layered Double Hydroxide Polymer Nanocomposites*, 1–76, Elsevier.

[4] Wei, J., Chu, X., & Li, S. (2019). Machine Learning in Materials Science. *Infomat*, 1(3), 338–358.

[5] Pal, M. (2005). Random Forest Classifier for Remote Sensing Classification. *International Journal of Remote Sensing*, 26(1), 217–222.

[6] R, S. M. (2004). Machine Learning Benchmarks and Random Forest Regression. *UCSF: Center for Bioinformatics and Molecular Biostatistics*.

[7] Maulud, D., & Abdulazeez, A. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Advanced Science and Technology Trends*, 1(2), 140–147.

[8] McDonald, G. C. (2009). Ridge Regression. *WIREs Computational Statistics*, 1(1).

[9] Awad, M., & Khanna, R. (2015). Support Vector Regression. In *Efficient Learning Machines*, 81–107, Apress, Berkeley, CA.

[10] Kang, M. (2005). Synthesis and Characterization of Layered Double Hydroxides by Precipitation Method.

[11] Lee, H. S., Kim, M., Kim, J., & Oh, S. (1998). Synthesis and Characterization of Layered Double Hydroxides by Precipitation Method. *Journal of the Korean Chemical Society*, 42(4), 432–442.

[12] Ahn, B. Y. (1994). X-ray Diffraction Principles and Applications. *Journal of the Korean Chemical Society*, 7(2), 1038–1067.

[13] Jeong, S., & Chun, J. (2010). Principle and Application of Scanning Electron Microscope. *Industrial Chemistry Prospect*, 13(1), 51–60.

[14] Severin, K. P. (2004). Energy Dispersive Spectrometry, 1–13, Springer Netherlands.

[15] Ul-Hamid, A. (2018). A Beginners' Guide to Scanning Electron Microscopy, 402, Springer International Publishing, Cham.

[16] Gaines, P. R. (2011). ICP Operations Guide: A Guide for Using ICP-OES and ICP-MS. *Inorganic Ventures*, 44, 4–8.

[17] Smith, E., & Dent, G. (2019). Modern Raman Spectroscopy: A Practical Approach, John Wiley & Sons.

[18] Haddad, P. R., & Jackson, P. E. (1990). Ion Chromatography, Elsevier.