Materials Horizons

Predicting Rare Earth Elements Concentration in Coal Ashes with Multi-Task Neural Networks

Journal:	*Materials Horizons*
Manuscript ID	MH-COM-09-2023-001491.R2
Article Type:	Communication
Date Submitted by the Author:	18-Dec-2023
Complete List of Authors:	Song, Yu; University of California Los Angeles, Zhao, Yifan; University of California Los Angeles Ginella, Alex; University of California Los Angeles Gallagher, Benjamin; Electric Power Research Institute Sant, Gaurav; University of California Los Angeles Bauchy, M.; University of California Los Angeles

We demonstrate a novel concept that addresses the pressing issue of Rare Earth Element (REE) extraction from waste coal ashes through multi-task machine learning. This approach allows a single machine learning model to simultaneously learn from test data measured for multiple REEs in coal ash samples. As a result, the model training process becomes significantly more efficient, even when working with a limited-sized REE dataset.

This novel multi-task machine learning framework is designed to predict REE concentrations in coal ashes efficiently and affordably. This concept sets itself apart from existing research by being the first to enable precise predictions of REE concentrations solely from simple X-ray fluorescence (XRF) measurements, eliminating the need for costly and advanced material characterizations.

Importantly, our innovation not only revolutionizes high-throughput screening for REE-bearing coal ashes but also has broader implications for the field of materials science. It effectively addresses the common challenge of dealing with sparse material datasets with multiple interrelated properties, making it a versatile tool for researchers in this domain. Our work thus contributes an additional layer of insight to the field of machine-learning-based materials science by presenting an efficient solution to enhance modeling capabilities when working with limited material datasets.

Full title

Predicting Rare Earth Elements Concentration in Coal Ashes with Multi-Task Neural Networks

Yu Song,[1,2]* Yifan Zhao,[1] Alex Ginella,[1] Benjamin Gallagher,[3] Gaurav Sant,[2,4,5,6] and Mathieu Bauchy [1,4]*

[1] Affiliation 1:
Physics of AmoRphous and Inorganic Solids Laboratory (PARISlab)
5731B Boelter Hall, Department of Civil and Environmental Engineering, University of California Los Angeles, CA 90095, USA.

[2] Affiliation 2:
Laboratory for the Chemistry of Construction Materials (LC^2)
5731J Boelter Hall, Department of Civil and Environmental Engineering, University of California Los Angeles, CA 90095, USA.

[3] Affiliation 3:
Electric Power Research Institute (EPRI)
3420 Hillview Avenue, Palo Alto, CA 94304, USA.

4 Affiliation 4:
Institute for Carbon Management (ICM), University of California, Los Angeles, CA, USA.

[5] Affiliation 5:
Department of Materials Science and Engineering, University of California, Los Angeles, CA, USA.

[6] Affiliation 6:
California Nanosystems Institute, University of California, Los Angeles, CA, USA.

* Correspondence: yusong@ucla.edu; bauchy@ucla.edu

Abstract: The increasing demand for rare earth elements (REEs) makes them a scarce strategic resource for technical developments. In that regard, harvesting REEs from coal ashes—a waste byproduct from coal power plants—offers an alternative solution to conventional ore-based extraction. However, this approach is bottlenecked by our ability to screen coal ashes bearing large concentrations of REEs from feedstocks—since measuring the REE content in ashes is a time-consuming and costly task requiring advanced analytical tools. Here, we propose a machine learning approach to predict the REE contents based on the bulk composition of coal ashes (which is easily measurable under the current testing protocol). We introduce a multi-task neural network that simultaneously predicts the contents of different REEs. Compared to the single-task model, this model exhibits notably improved accuracy and reduced sensitivity to noise. Further model analyses reveal key data patterns for screening coal ashes with high REE concentrations. Additionally, we showcase the utilization of transfer learning to improve the adaptability of our model to coal ashes from a distinct source.

Teaser: With machine learning, high-throughput screening of REE-bearing coal ashes can be fulfilled based on a simple measurement.

Keywords: Rare Earth Element; Machine Learning; Screening; Coal Ash

1. Introduction

Rare earth elements (REEs) are a collective term for 17 elements that are commonly found in a variety of minerals, but typically in low concentrations. The members of REEs include 15 elements of the lanthanoid series (La, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu), yttrium (Y), and scandium (Sc). While they are not abundantly present as compared to other elements that are commonly seen in ore deposits (*1*), these elements are playing an increasingly vital role in many high-tech industries that are closely related to clean energy (e.g. high-efficiency lighting, advanced fuel cell systems, wind energy, and electric mobility) (*2*–*5*), along with many other key fields (e.g., permanent magnet motors and generators, semiconductor chips, optics, catalysts, and defense technologies) (*6*–*10*). Given the rapidly growing demand for REEs from various sectors in recent decades, the supply chain of these materials is currently under heavy pressure, and this pressure is anticipated to continue increasing at a 5-9% annual growth rate globally in the next three decades (*3*). For instance, according to the U.S. Geological Survey, the country's REE consumption was fully relied on imports in 2019 (*3*, *11*); in the face of the short supply, the White House has reiterated the strategic importance of developing REE-recovering technologies in form of an executive order in early 2022 (*12*).

Given the low supply and high demand for REEs from ore deposits in the predictable future, many recent studies have paid special attention to identifying new sources of REEs. Among the various available sources, coal ash (the burning residue of coal from power plants) has been considered as an important potential source of REEs (*13*–*18*). In that regard, the Department of Energy (DOE) has allocated $140 million to recover REEs from coal ashes (*19*). The large global reserves of coal have attracted numerous research efforts to extract REEs from coal ashes, whose worldwide average REE concentrations are estimated at around 400 ppm (by mass) (*18*). In addition, recovering REEs from coal ashes has several potential advantages over the ore-based extraction (*14*), which include obtaining higher concentrations of heavy and critical REEs (*18*) (which are much lower in supply, higher in price, and projected for a greater demand), incurring negligible mining cost (since coal ash is a byproduct from the coal production), and precluding the need of handling radionuclide hazards form the ore deposits. In recent years, a large number of studies have been published to promote the concept of extracting REEs from coal ash, and the relevant topics broadly revolve around (i) geological distributions, depositional settings, and resource characteristics of coal deposits containing REEs (*17*, *18*, *20*, *21*), (ii) novel process

techniques that separate REEs from coal ashes (*22*–*24*), and (iii) advanced methods detecting the concentration of REEs in coal ashes, such as inductively coupled plasma mass spectrometry (ICP-MS) (*25*, *26*), inductively coupled plasma optical emission spectrometry (ICP-OES) (*27*), SHRIMP-RG ion microprobe (*28*), and laser-induced breakdown spectroscopy (LIBS) (*29*).

Historically, coal ashes have commonly been used as a cement replacement in concrete (as supplementary cementitious material, SCM) (*30*–*32*). To minimize the negative impact of cement replacement on concrete performance (e.g., strength reduction), certain restrictions on the physical properties (e.g., particle fineness) and chemical composition (e.g., CaO and unburnt carbon contents) of the coal ashes are typically imposed during recycling (*33*). As a consequence, this approach results in a considerable volume of off-specification coal ashes among those that are currently deposited in impoundments or landfills (*34*). Legacy coal ashes that have already been deposited in impoundments may pose a risk of environmental damage to their surrounding groundwater system due to chemical leaching (*35*, *36*). In this sense, recovering REEs from legacy ashes further offers a potential environmental benefit to mitigate the pollution issue associated with coal ashes deposition.

For both freshly produced coal ashes and those deposited in impoundments, a batch-wise screening of the coal ashes is necessary for an effective REE extraction. This is because the REE concentration from a given coal source is affected by the combustion protocols, processing techniques, as well as storage conditions (*13*, *14*, *37*). Thus, to enable an efficient REE extraction, it is of special significance to develop accurate, rapid, and high-throughput screening methodologies that can identify the coal ashes presenting the highest potential for REE extraction, that is, the ashes featuring the largest REE content. From the production's perspective, almost all the characterization techniques (e.g., ICP-MS or ICP-OES) for determining the REE concentration in coal ash are time and cost-prohibitive for real-time ash screening. In contrast, X-ray fluorescence (XRF) is a fast, convenient, and economical measurement that has been widely adopted for testing the composition of coal ashes. However, XRF is only capable of measuring the bulk contents of major elements (i.e., elements showing elevated concentration) and does not offer the level of accuracy that is required to detect REEs. Nevertheless, previous studies have suggested the existence of a certain correlation between the total REE content and the bulk composition of coal ashes (*28*, *38*, *39*). For instance, REEs tend to concentrate more in the amorphous aluminosilicate phase in coal ashes.

To enable accurate, rapid, and high-throughput coal ash screening for REE extraction, it is imperative to establish a robust mapping that bridges the easily-measurable bulk XRF measurement to the total REE content. However, existing physical and chemical models are presently unable to support such a compositional mapping. In that regard, data-driven machine learning analysis offers a promising pathway to accomplish this mapping without the need for explicit knowledge regarding the nature of the compositional mapping. A major strength of machine learning is that it can uncover complex, non-additive, and nonlinear patterns embedded in the data, whereby the mapping function between variables can be revealed without the presence of explicit knowledge (*40*). In fact, over the decade, machine learning techniques have enabled many critical advances in materials property prediction, material design, advanced characterization, new functional material discovery, high-efficiency computational methods (*41*–*49*), as well as REEs-related topics (*50*, *51*). In particular, a few recent studies have demonstrated the feasibility of using machine learning analysis to rapidly screen fly ashes (*52*, *53*), or REE-bear coal material (*53*). However, to the best of our knowledge, no machine learning approach has thus far been attempted for screening REE-bearing coal ashes.

In this paper, we explore the potential of using machine learning to predict the total REE content in coal ashes based on the sole knowledge of the easily-measurable bulk oxide contents. This kind of input information is readily available under the prevailing coal ash testing protocols, via XRF measurement. As such, this approach enables a fast, ready-to-use screening technique for coal ashes that can be applied at large-scale production/reclamation job sites, where batch-wise feasibility of REE extraction can be determined in real-time, before conducting subsequent characterization at a finer level. It is worth noting here that our data-driven approach aims to investigate the feasibility of inferring the REE content based on the presence of the major oxides. While the modeling work takes no assumption that the REEs are directly related to the presence of bulk oxides, they may present as two correlated consequences under the same hidden physics law (e.g., thermal history). The machine learning model is trained upon a coal ash REE dataset comprising 99 representative samples. To address the limited size of the dataset, a series of data processing techniques are implemented to boost the accuracy of our machine learning models. We adopt neural networks using as input the bulk composition of merely six oxides to predict as output the total REE content in coal ashes.

As the key advancement of this study, we introduce multi-task neural network modeling to enable a simultaneous prediction of the total and the individual REE contents. While multi-task machine learning approaches have been proposed for years (*54*, *55*), their potential for addressing material-related

problems has not been sufficiently explored. Herein, the use of a multi-task model is well-justified because certain REEs are known to co-occur in coal deposits (*56*). In addition to the reduced time and computational demands for training multiple single-task models, multi-task modeling provides significant merits, such as implicit data augmentation and regularization, that can enhance the accuracy of model predictions (*57*). The results indicate that, as compared to all the conventional single-task models considered in this study, our multi-task model achieves higher accuracy and lower sensitivity to input noise in predicting the total REE content. Further, our model analyses highlight the opportunity of extracting REEs from reclaimed coal ashes, which, on average, tend to exhibit higher REE concentrations than raw coal ashes.

2. Experimental procedures

2.1 The coal ash dataset

The coal ash dataset adopted herein was sourced by the Electric Power Research Institute (EPRI), based on the experimental work done at National Energy Technology Laboratory (NETL). This dataset comprises 99 entries, which include 18 coal samples and 81 coal ash samples. These samples were originally investigated to represent the systematic variations in coal ash compositions across the United States of America, which cover four coal ranks (i.e., bituminous, sub-bituminous, lignite, and blend), three types of coal boilers (i.e., pulverized, cyclone, integrated coal gasification combined cycle), ten collection origins (e.g., fly ash, bottom ash, pond ash, and landfill ash), and nine different coal basins. All the samples were measured following the procedure described in DOE/NETL-2016/1794 report (*58*). In detail, for either the raw coal or coal ash samples, the moisture and organic compounds in the sample were first removed by calcination in a high-temperature furnace at 1100°C. After that, the burning residue was ground into fine powders that are less than 74 μm (200 mesh) in diameter, fused with lithium metaborate, and further fully dissolved in concentrated acid solutions. The concentration of each element in the sample was then measured using inductively coupled plasma mass spectrometry (ICP-MS), and the results were reported in terms of the mass fraction of the individual elements in the raw sample (i.e., either coal or coal ash). For each sample, the raw dataset reports the mass contents for a total of 51 major and trace elements, the latter of which include all the 16 REEs except for promethium (Pm). To filter out the effect of the molar mass of each element from the raw ICP data, we convert all the mass contents of the 51 elements measured from the calcinated samples into the molar contents of their corresponding oxides. The molar contents of those oxides are then renormalized based on the total oxide quantity in each sample.

Since the goal of this study is to predict the content of REEs in ashes based on that of the non-REEs, the 51 oxides are divided into two groups (i.e., 35 non-REE and 16 REE oxides). With the specific aim of rapid ash screening in mind, we limit the input information for our machine learning model (hereafter, input feature) from the 35 non-REE oxides down to 10 major oxides. These oxides comprise more than 98% of the oxides in each sample (see Table 1), and, importantly, they are typically measurable in coal ash XRF characterization (*31*, *59*). Although there is a difference between the ICP and XRF tests in that sulfur is not reported in the former, it is reasonable to omit SO_3 as it has a low concentration in coal ashes. Hence, the model inputs adopted herein can be easily obtained from the XRF measurement—whereby the screening of high REE-bearing coal ashes can be greatly facilitated in industrial applications. A subset of 10 samples in the curated dataset is provided in in Supplementary Materials (Table S1).

2.2 Techniques for addressing the sparse nature of the coal ash dataset

To address the sparsity of the coal ash dataset, we first implement a recursive feature elimination (RFE) feature analysis to scrutinize the change in the accuracy of our neural network model when progressively removing the least influential input features (*60*). The core idea of RFE is that, in each iteration of the feature removal, the change in model accuracy associated with the removal of each individual left in the feature pool is first evaluated, and then only the feature associated with the least influence on the model accuracy (i.e., the least informative feature) is dropped. Following such a step, the feature pool can be progressively reduced to only the most valuable features. This allows us to determine whether some of the ten input features in the coal ash dataset can be removed from consideration. To ensure a reliable RFE analysis, we evaluate the model accuracy based on 100 different extractions of the validation set samples in each iteration before removing a feature (while the model is only trained with the training set samples).

We also adopt the stratified sampling technique for splitting the coal ash dataset into training, validation, and test sets. For handling small datasets in machine learning analysis, a common yet easily overlooked problem is that a random split of the dataset cannot guarantee a good statistical representation of samples being extracted (*41*). This is because the samples in a small dataset tend to reside sparsely in the high-dimensional space, while a random extraction often only covers part of the space. As a result, the direction of model optimization (as guided by the training set samples) may be misled into a local optimum in the feature space, instead of toward the direction of a global optimum. To overcome this problem, the stratified sampling approach divides the raw dataset into several subsets based on the distribution of the output target (i.e., total REE content) and then extracts the training samples from each

of the subsets by proportion. Doing so helps to ensure a more robust evaluation of the true performance of the trained model (and, in turn, higher efficiency of the subsequent model training). Following our previous study (*52*), we divide the coal ash dataset into five portions and implement the stratified sample accordingly.

2.3 Neural networks modeling

We adopt artificial neural networks for predicting the output target (i.e., the total REE content) based on the input features (i.e., the contents of common oxides that can be easily measured by XRF). Here, we specifically adopt neural networks as a regression model because of their good flexibility in handling different types of datasets and interpolating complex relationships between the input features and output target (*61*). This approach has been previously adopted for studying many glass-based materials (*41*, *43*, *44*, *48*), including coal ashes (*62*, *63*). Herein, the modeling process follows one of our previous studies that also used neural networks to analyze the chemical reactivity of coal ashes (*52*), wherein the basics of neural networks were thoroughly reviewed. The following contents focus on the core concepts of neural networks and the modeling details exclusive to the investigation of REE prediction. The modeling work, along with the analyses, in this study is conducted in Spyder IDE (version 5.4.3) with Python (version 3.10.10), and the specific program libraries are further detailed in the text. All the artificial neural networks presented herein are implemented using the PyTorch deep learning library (version 2.0.0) (*64*). The other machine learning models investigated involved in this study are built using the scikit-learn (version 1.3.0) and XGBoost (version 1.7.6) libraries, as further detailed in Sec. 3.2 (*40*, *65*). The core code of the modeling work is available on GitHub (see Sec. S5).

Figure 2a depicts the architecture of one of our neural network models (in this case, a single-task model with only one output). Within this architecture, each artificial neuron is a computational unit that handles a basic learning task, while the whole ensemble of these computational units is capable of mapping the complex relationship from the input to the output. Considering the limited size of the REE dataset while ensuring sufficient learning ability of the model, we adopt a relatively simple network structure of two layers of artificial neurons (which are known as hidden layers) for computing the relationship between input features (i.e., XRF compositions) and the output target (i.e., total REE content). An illustration of our selection of the two-layer model is provided in Sec. S2.

In detail, the artificial neurons in the first hidden layer are designed to extract low-level features directly from the inputs of the XRF composition, while those in the second layer are responsible for

220 further synthesizing the low-level features to obtain a final prediction (of the REE content in the coal ash).
221 After each of the two hidden layers, we further add two functional layers, namely, batch normalization and
222 nonlinear activation. The batch-normalization layer improves our model's learning efficiency by avoiding
223 any significant elongation of the loss landscape along a given direction (note that we use a batch of ten
224 samples in each iteration for the model training) (*66*). The activation layer—here, Rectified Linear Unit
225 (ReLU)—enables the model to learn nonlinear relationships for mapping the inputs to the output. In
226 addition, we adopt the Adaptive Moment Estimation (Adam) algorithm to optimize the artificial neurons
227 throughout model training. To ensure sufficient model training, we further fix the training epoch number
228 to a relatively large value of 500 (wherein each epoch corresponds to one iteration of using all the samples
229 for training the model). This number of epochs is found to be large enough to ensure a convergence of the
230 loss function (herein, mean squared error, MSE).

231 In detail, the entire modeling work is divided into two phases, namely, model training and testing.
232 The training phase consists in optimizing the parameters of the neural network model to improve its
233 prediction accuracy on a subset of samples in the coal ash dataset (i.e., training set), while the testing
234 phase is used to evaluate the actual performance of the trained model on the other subset of samples that
235 are kept hidden to the model during the training phase (i.e., test set). Herein, we allocate 80% of the
236 samples to the training set (wherein 20% of these samples are used as the validation set) and the
237 remaining 20% to the test set. We primarily use the coefficient of determination (R^2) to evaluate the model's
238 accuracy and further consider MAPE.

239 In the model training phase, we compare the model performance under the choices of three
240 hyperparameters, namely, (i) number of artificial neurons, (ii) learning rate, which controls the rate for
241 updating the parameters of the artificial neurons in each training iteration, and (iii) weight decay, which
242 adjusts the level of model fitting based on L_2 regularization. The hyperparameter selection is a critical step
243 to determine the optimal complexity of the neural network model, so that it is neither too complex for
244 the present coal ash dataset (to mitigate the risk of overfitting for such a small dataset), nor too simple to
245 capture the true mapping function involved with the REE prediction (to avoid underfitting). Note that
246 those hyperparameters need to be predetermined before the training of the machine learning model and,
247 thereafter, are kept unchanged during model training and testing. For technical details about those
248 hyperparameters, the reader is invited to refer to the webpage of the PyTorch library and other sources
249 such as Ref. (*67*).

We carry out the hyperparameter selection based on a grid search, by systematically varying the three hyperparameters within their respective reasonable ranges. The number of artificial neurons is varied from 2 to 32 (with an incremental factor of 2) in each layer; learning rate is varied from 10^{-5} to 10^{0} (with an incremental factor of $10^{0.33}$); and weight decay is varied from 10^{-6} to 10^{-1} (with an incremental factor of $10^{0.33}$). Following common practices in training machine learning models (*43*, *68*), we split a small portion (20%) from the training set to validate the model performance based on the selected hyperparameters. Based on the model accuracy observed from the validation samples, we determine the optimal hyperparameters as (i) 20 and 6 neurons in the two hidden layers, (2) $10^{-3.33}$ for learning rate, and 10^{-2} for the weight decay. Based on additional model comparisons, further increasing the model complexity, such as using more hidden layers, artificial neurons, or additional functional layers, does not result in notable improvements in the model performance. As a demonstration, the variation of model accuracy with different hidden layers is shown in Supplementary Materials (Sec. S2).

Once the hyperparameters are optimized and fixed, we move on to the testing phase to evaluate the actual prediction accuracy of the fully-trained model. To this end, we first retrain the neural network model using all the 80% training samples (including the validation samples) with the optimal hyperparameters—which further boosts the model accuracy. We then evaluate the prediction accuracy of the retrained model by using the testing samples that are never exposed to model training. To avoid the randomness associated with neural networks, the evaluation is conducted based on 30 repetitions of the model training.

2.4 Multi-task neural networks

Since, in the present case, several outputs are simultaneously available (i.e., the content of each type of REE) and these outputs exhibit some level of correlation, we then investigate whether using a multi-task neural network model could enhance the accuracy of the prediction of total REE content—as compared to a traditional single-task neural network that only predicts one output at a time. Multi-task, or multi-output, learning is an emerging approach in the field of machine learning (*69*, *55*, *54*), which has been drawing increasing attention in recent years (*70*, *71*). The basic idea of this approach is to use a single machine learning model to simultaneously predict multiple targets that share the same input features. By ingesting the entire dataset, the multi-task model can transfer the feature mapping learned from one target to the others. Such a synergic approach has been recently demonstrated to be highly effective in studying polymers, wherein the different material properties are also highly correlated (*72*).

For the present coal ash dataset, the rationale for applying multi-output learning lies in the facts that (i) the total REE content is the sum of the 16 individual ones and (ii) due to the presence of REE-bearing phases in coal ashes, a higher content of a single REE tends to suggest high contents of other REEs (and the total REE content thereof) (*56*). Figure 2b illustrates the architecture of the multi-task neural network model that is adopted in this study. To allow a fair comparison, this multi-task model is designed to have the same structure as the single-task model (Fig. 2a), with the only difference being the size of the output layer. Herein the output layer consists of a total of 17 targets, which include the total REE content and the individual contents of the 16 REEs (in their oxide forms). During the modeling process, the artificial neurons in this architecture are trained to attain the highest average accuracy over all 17 targets. To bolster the prediction of the total REE content, we further optimize the multi-task model by adjusting the weights associated with each target in the cost function, wherein we use the mean value of each target as weight.

2.5 Evaluation of noise sensitivity of the single and multi-task neural networks

After investigating the model accuracy, we further compare the influence of input noise on the single and multi-task neural networks. This comparison is carried out on the entire dataset and for each of the 30 model repetitions. Herein, we add two levels of Gaussian noise to each input feature. Based on the typical uncertainty of the XRF measurement (*73*, *74*), the standard deviation of the individual input is set to 2% and 4% of the original value, respectively. The added noise is intended to simulate the uncertainty of the XRF measurement for the major elements in the high-throughput screening of coal ashes. For both the optimized single and multi-task neural network models, we record the change in the model prediction for the total REE content when subjected to the noisy inputs. This allows us to investigate the distribution of the variation of the model predictions between those two models. To facilitate the comparison of the two models, the resultant distributions are further fitted with normal distributions (see Fig. 3).

2.6 Interpretation of the trained machine learning model

To interpret the patterns that are learned by the optimized multi-task model, we implement two independent model analyses, namely, a SHapley Additive exPlanations (SHAP) analysis (*75*) and a feature effect analysis. The SHAP analysis evaluates the trained model based on the Shapley value (this is, the marginal contribution of each input feature to the output target) (*52*, *76*). In particular, a major advantage of the SHAP analysis is that this approach considers the global influence of multiple input features as a

whole. As such, the SHAP analysis helps to capture the features associated with a significant joint impact, while those features may appear weakly influential when being evaluated individually.

In comparison, the feature effect analysis focuses on quantifying the change of the model prediction upon the variation of single input features (i.e., only one feature is jittered each time) (*77*). To perform the feature effect analysis, we first identify a reference composition based on the median value of each input feature, which serves as a representative baseline for the different samples in the coal ash dataset. Starting with this reference composition, we investigate the effect of each XRF content input on the total REE content. To avoid interrogating the response of the model to unrealistic coal ash compositions, we only jitter the feature of interest within its 20th-to-80th percentile range in the coal ash dataset (see Table 1) and maintain all the other features fixed. We then track the evolution of the output(s) predicted as a function of the value of the varying feature. After looping through all the input features, we compare their effects collectively in a single plot.

3. Results

3.1 Features selection

To investigate the REE concentration in coal ashes, we curate a dataset comprising representative 99 coal ash samples, as detailed in Sec. 2.1. We select 10 major oxides to represent the bulk composition of each sample (i.e., model inputs), while the target of the model prediction is the molar contents of 16 REEs (i.e., model outputs; note that Pm is not considered as this element is practically non-existent in nature). We further use their sum as the ultimate prediction target (i.e., total REE content). The statistics of this dataset are summarized in Table 1, and the correlation between the 16 individual REEs and the total REE content is displayed in Fig. 1a. Here, we observe some strong correlations between the contents of different REEs, which agree with the general idea that different types of REEs tend to be found within same chemical phases in coal ashes (*56*). More importantly, several experimental studies observed certain correlations between the presence of REEs in coal ashes and the bulk composition of some major elements (29, 39, 40). These observations provide a ground for mapping the bulk chemical composition of coal ashes to REE contents with machine learning, which is known to excel in capturing implicit data patterns without explicit knowledge (*41*).

Table 1: **Statistical distribution of the coal ash samples considered by the present machine learning analysis.** Note that the content of each oxide presented here is normalized based on those of all oxides measured using inductively coupled plasma mass spectrometry (ICP-MS, see text).

Statistics		Oxide fraction [molar%]											Total REE
		SiO_2	Al_2O_3	CaO	Fe_2O_3	MgO	Na_2O	K_2O	TiO_2	P_2O_5	MnO	Sum	[ppm, by molar]
Percentile	0	33.3	8.0	1.2	1.5	0.6	0.2	0.1	0.5	0.0	0.0	98.3	80
	20	44.4	10.9	2.1	2.6	2.1	0.3	0.4	0.9	0.1	0.0	99.5	117
	50	63.0	13.7	6.0	4.9	2.7	0.7	1.2	1.1	0.2	0.0	99.6	157
	80	67.9	18.5	27.2	8.2	7.4	1.3	1.7	1.3	0.4	0.0	99.8	246
	100	79.3	21.3	37.7	12.8	14.1	13.3	2.3	2.6	0.9	0.8	99.9	293
Mean		58.1	14.3	13.7	5.5	4.5	1.0	1.1	1.1	0.2	0.1	99.6	170
Standard deviation		11.6	3.7	12.5	3.0	3.0	1.4	0.6	0.3	0.2	0.1	0.2	60

A major challenge of the machine learning analysis conducted in this study arises from the fact that the curated coal ash dataset has a relatively high dimensionality of input features (i.e., 10 by default; see Table 1) as compared to its small number of datapoints (99 samples), thereby resulting in a very sparse dataset. This issue makes it difficult to train a machine learning model efficiently—typically refers to as the "curse of dimensionality" (*78*). In that regard, the main concern arises from the fact that some of the ten input features may exhibit a weak correlation to the RRE content, thereby causing their contributions to the model prediction to be overwhelmed by noise (e.g., due to the systematic or random errors of the measurement on the oxide contents). As such, some of the weak features may actually impede the learning process of the machine learning model and, consequently, result in lower prediction accuracy and/or a higher instability for the trained neural network model.

To address the data sparsity challenge, we first focus on reducing the dimensionality of the input features (i.e., the major oxides summarized in Table 1). To this end, we implement a recursive feature elimination (RFE) analysis that aims at determining the optimal combination of the input features by progressively removing the least influential ones to the model prediction (see Sec. 2.2) (*60*). This analysis allows us to analyze the contribution of each feature to the prediction accuracy of our neural network model and gradually exclude the non-informative oxides from the consideration. To carry out this analysis, we use a two-layer neural network as the base model, which takes as input the concentrations of the major oxides and predicts as an output the total REE content. This base model is described in Sec. 2.3. The features are recursively selected based on the model accuracy on the validation samples, with 100 random repetitions. Note that, different from some other unsupervised approaches that simply aim to reduce the feature dimensionality such as principal component analysis (PCA), RFE is a supervised approach since

this approach specifically selects the features that are the most informative to predict the output target (herein, the REE content).

The results of the RFE analysis are given in Figs. 1b and 1c. To ease the discussion, the prediction error displayed in Fig. 1b is normalized based on the mean absolute percentage error (MAPE) value achieved when all the ten features are used for training the model (i.e., when the model is trained with the maximum amount of information). With the features being iteratively added to the model, we initially observe a decline in the prediction error (and, hence, an improvement of the model accuracy) up to using six features. However, the error then flattens out when more than six features are used in the model. Eventually, we observe a slight increase in the prediction error. This suggests that these additional features do not provide meaningful information for improving the model accuracy and, rather, unnecessarily increase the risk of overfitting by increasing the model complexity and level of noise that the model is exposed to. By repeating the RFE analysis 100 times, we then determine the probability for each input feature (i.e., each oxide) to be selected as a function of the number of features being considered, that is, as a function of the RFE iteration step (see Fig. 1c). We observe that the oxides that are selected early on (i.e., even when the number of considered inputs is small) tend to continue to be selected after adding more features. This suggests that this feature selection analysis yields a stable selection of the oxides that are the most informative to the model. Based on Fig. 1c, the molar contents of Al_2O_3, Fe_2O_3, CaO, P_2O_5, TiO_2, and MgO are determined to be more influential than the remaining four ones to infer the presence (or absence) of the REE-bearing phases. However, the results in Fig. 1c should not be interpreted in a way such that the unselected features are insignificant—rather, they are simply not contributing to further boost the prediction accuracy (e.g., due to a high correlation with the selected features).

Therefore, these six oxides are selected as input features for training the machine learning models presented in this study. For additional insight, the correlation between the contents of the six oxides and individual REEs is detailed in Supplementary Materials (Sec. S3). It is apparent that the presence of REEs exhibits a degree of alignment with Al_2O_3, although the correlations with the other oxides are notably non-linear. Additionally, the correlation to an REE can vary across the oxides. These factors are expected to introduce certain complexities for predicting REE content based on oxide content.

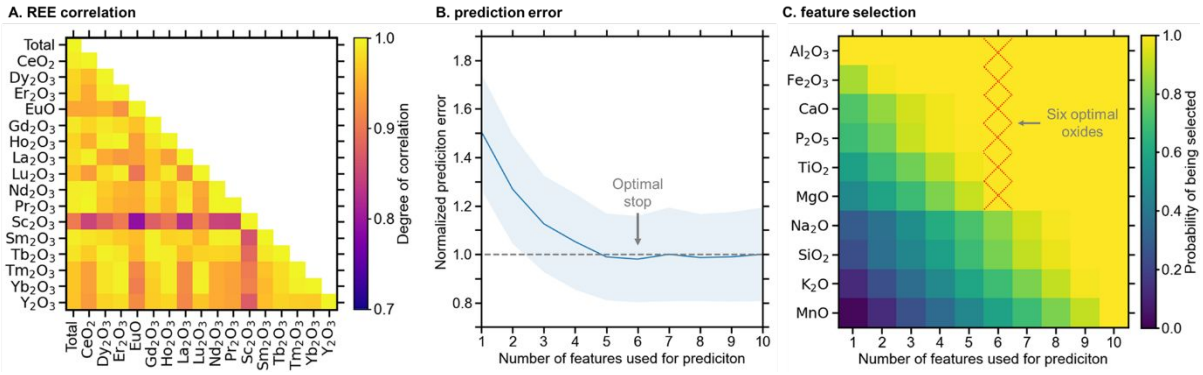

Figure 1: **Feature selection for the coal ash dataset. (A)** Correlation heat map for the molar contents of the REEs (based on their corresponding oxide forms). Results of the recursive feature elimination (RFE) analysis: **(B)** variation of the model prediction error as a function of the number of considered input features, and **(C)** probability for each input feature (i.e., major oxides in coal ashes) to be selected at each step of the RFE analysis. For **(B)**, the prediction error is normalized based on the mean absolute percentage error (MAPE) when all ten features are used, and the shadow indicates the standard deviation from 100 repetitions (see Sec. 2.2). For **(C)**, the cross marks indicate the optimal set of the six features to be used as model inputs.

3.2 Single-task vs. multi-task neural network models

Using the six oxide features selected from the RFE analysis, we then evaluate the accuracy of our single-task neural network model (see Fig. 2a) in predicting the total REE content. This neural network model is further detailed in Sec. 2.3. As a benchmark, we also investigate the prediction accuracy obtained by a multivariate linear regression model, which is arguably the most basic machine learning model (*42*). The training and test accuracy of the two models are compared in Table 2. Note that, other single-task machine learning models (i.e., random forest, SVM, XGboost) have also been compared during our experiential work; however, those models did not exhibit superior performance than the single-task neural network model discussed herein. As a reference, a comparison of the different single-task models is available in Supplementary Materials (Sec. S4).

To ensure a fair comparison, identical train-test splits are used for the neural network and multivariate linear regression models—see Figs. 2c and 2d for a visual comparison of the two models. From Table 2, we first observe that stratification—a sampling technique implemented herein to address the data sparsity (see Sec. 2.2)—offers an evident improvement in the accuracy of the single-task neural network, as well as a reduction in the associated standard deviation. As expected, stratification meaningfully improves the robustness of the train-test split for this small coal ash dataset, in line with our previous results (*52*, *79*).

Importantly, we find that, with the test set MAPE being lower than 10%, the single-task neural network model outperforms the linear model by a considerable margin, both in terms of model accuracy and stability (i.e., low standard deviation). Altogether, it is remarkable that the model achieves this accuracy based on the sole knowledge of the contents of the six major oxides. This confirms the potential of using the knowledge of the bulk oxide contents (e.g., via XRF) to conduct the rapid and high-throughput screening of REE content in coal ashes.

Table 2: **Comparison of the accuracy in predicting the total REE content for the two machine learning models considered herein.** The results are reported based on the coefficient of determination, R^2, and mean absolute percentage error, MAPE; and the mean and standard deviation are obtained from 30 repetitions of the model training.

		Multivariate linear regression		Single-task neural net	
Stratification		No	Yes	No	Yes
Training set	R^2	0.880 ± 0.014	0.876 ± 0.010	0.904 ± 0.041	0.940 ± 0.019
	MAPE [%]	9.6 ± 0.1	9.8 ± 0.5	8.5 ± 1.7	6.7 ± 1.1
Test set	R^2	**0.804 ± 0.087**	**0.847 ± 0.060**	**0.819 ± 0.119**	**0.875 ± 0.037**
	MAPE [%]	**11.1 ± 2.1**	**10.2 ± 2.0**	**10.8 ± 3.6**	**9.5 ± 1.3**

Given the high correlations between the individual REEs (Fig. 1a), we further explore the potential of using a multiple-task neural network (see. Fig. 2b) for simultaneously predicting the total REE content and the contents of the 16 individual REEs (i.e., with a total of 17 distinct outputs). The rationale and additional details about this multi-task model are provided in Sec. 2.4. To ensure a fair comparison between the single- and multi-task neural networks, we maintain the same network structure (except for the output layer) that is used for the optimized single-task neural network. For multi-task neural networks, a key parameter is the weights that are attributed to the prediction loss terms associated with each output target, which carry the attention that a multi-task model pays to optimize the different targets. By default, all these weights are equal to each other. However, it is often preferable to assign larger weights to the most important output targets (such that they can be given more attention during the model training) (*55*, *80*). In detail, the optimal weight assignment is contingent on the specific nature of the dataset and the goal of the machine learning analysis (*70*, *71*).

Here, we compare two multi-task neural network models. The first one is trained using equal weights. For the second one, we weigh the 17 targets based on their respective mean contents in the coal ash dataset, as a way to place more emphasis on the REEs that are the most predominant in the ashes and, *a fortiori*, on the total REE content. Figure 2e summarizes the prediction accuracy on all 17 targets,

as achieved by the two multi-task models, based on the same train-test splits as those used for the single-task neural network. As a key result, we find that the two multi-task models both exhibit a notably enhanced accuracy as compared to that achieved by the single-task model (see the accuracy comparison between all the models in Figs. 2f and 2g). In detail, we then note that the model with equal weights tends to accurately predict all the targets, with an average MAPE of approximately 10%. In particular, it achieves a higher test accuracy and a lower standard deviation in predicting the total REE content than the single-task models (see Table 2), where the average MAPE decreases from 9.5% to 9.0%. Remarkably, the prediction accuracy scores shown in Fig. 2e are achieved by a single model, whereas predicting these 17 outputs would have required 17 distinct neural networks if using conventional single-task models. Importantly, even when using the same weight for all the input features, the multi-task approach outperforms the single-task approach in predicting the total REE content. This broadly highlights the advantages of the multi-task approach (*57*), as further discussed in Sec. 4.1.

Next, as also seen in Fig. 2e, the weight-optimized multi-task model achieves even higher accuracy and smaller deviation in predicting the total REE as compared with the equal-weight model. However, this comes at the expense of a lower accuracy in predicting a few individual REEs, such as Ho, Hg, Tb, and Tm. This is not surprising since these REEs show the lowest average contents in the ashes considered herein (see Table 1) and, hence, have the lowest contribution to the loss function used in the weight-optimized model (which is accessible on GitHub as a reference, see Sec. S5). It should also be noted that, due to their lower concentration, these oxides contents are more challenging to quantify using ICP—their measurements are expected to be noisier than the oxides that exhibit higher concentrations. Nevertheless, we observe that the standard deviations associated with each target are consistently smaller when compared with the equal-weight model, which suggests that the weight-optimized model is intrinsically more stable. Overall, these results suggest that tuning the weight associated with each input feature can effectively increase the robustness of the multi-task model.

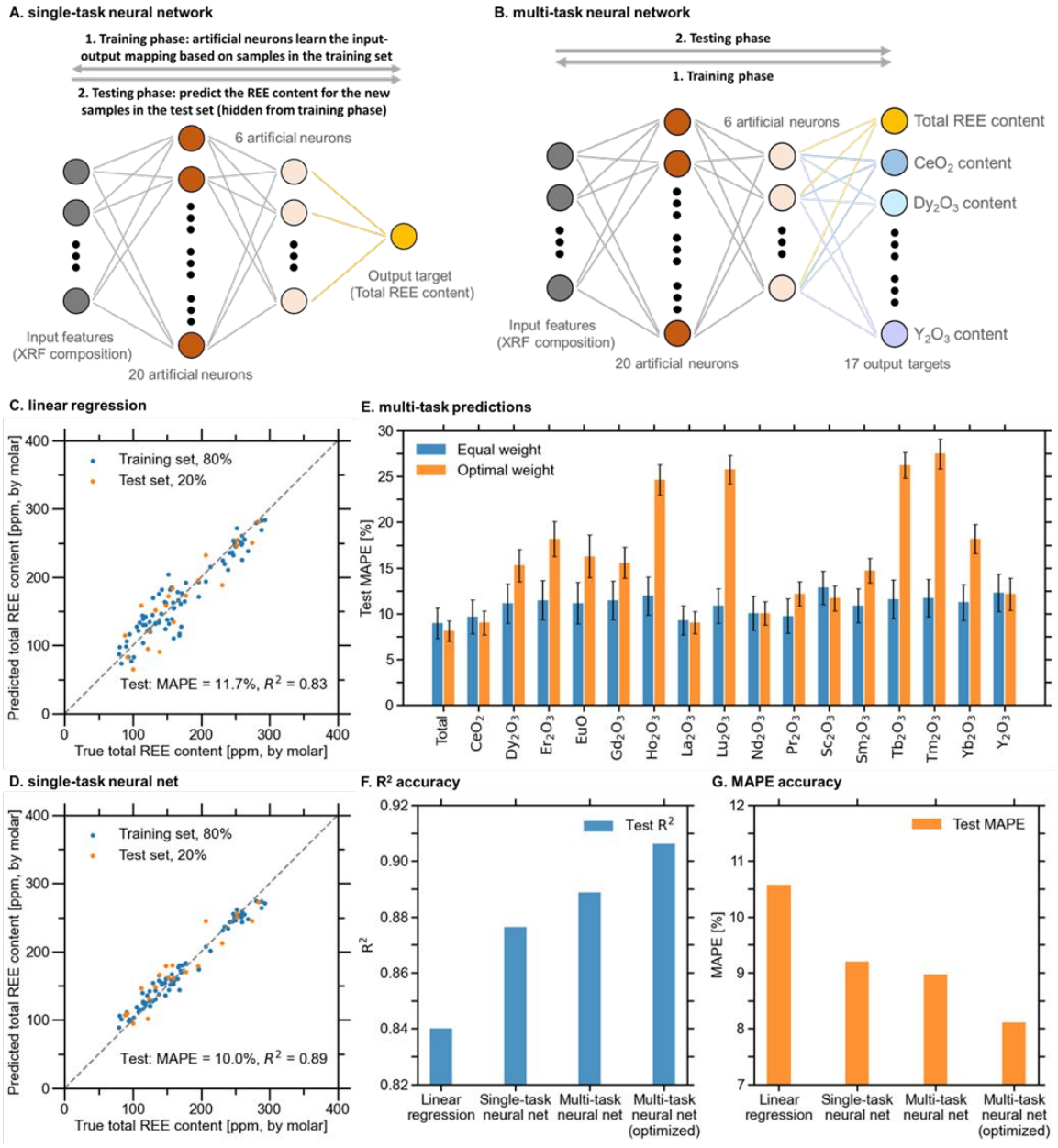

Figure 2: **Comparison of the prediction accuracy of the different models**. Illustration of the architecture of **(A)** the single-task neural network, and **(B)** the multi-task neural network, wherein both models contain the same structure of two hidden layers, while the output layer of the multi-task model simultaneously predicts the contents of 16 individual REEs and the total REE content (see Sec. 2.4 for details). Predicted versus measured total REE content, as obtained by **(C)** the multivariate linear regression model, and **(D)** the single-task neural network model, wherein the $y = x$ dashed line indicates the perfect agreement. **(E)** Comparison of the test accuracy between the two multi-task neural network models. Overall comparison of the test accuracy of the four major models in predicting the total REE content in coal ash, based on **(F)** coefficient of determination, R^2, and **(G)** mean absolute percentage error, MAPE. For **(E)** to **(G)**, the results are averaged over the 30 model repetitions of training.

The results above suggest that, by leveraging the information from multiple prediction targets, the multi-task model exhibits not only higher accuracy but also less variation than the single-task model in predicting the total REE content. A follow-up investigation is carried out to assess the sensitivity of the multi-task model to the noise from the model input. This is pertinent to the high-throughput screening of coal ashes, as XRF measurements are usually subject to variations (e.g. relative standard deviations of 1 to 5% for major elements in coal ashes) (*73*, *74*). Here, we introduce 2% and 4% Gaussian noise into each of the model inputs (see details in Sec. 2.5). The variations of the total REE prediction are shown in Fig. 3. While the variations of the model prediction are normally distributed, the multi-task model yields notably higher peaks and narrower spreads of the prediction variations at both the noise levels. These results demonstrate that the multi-task model does exhibit improved resistance to the input noise. Further discussion about this improvement is provided in Sec. 4.1.

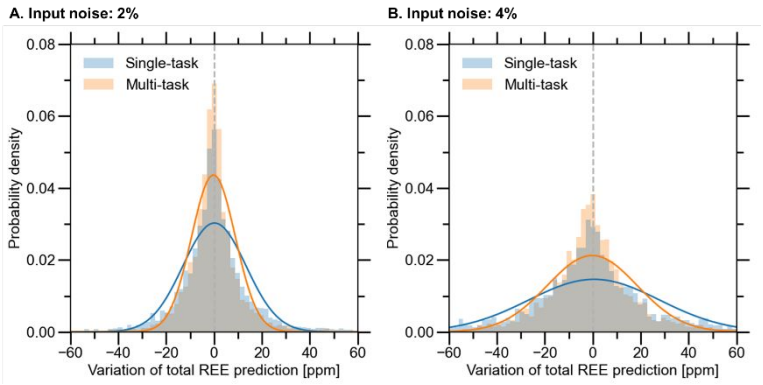

Figure 3: **Comparison of sensitivity to input noise between single- and multi-task neural networks for predicting the total REE content under different levels of noise.** Input noise levels are set at **(A)** 2%, and **(B)** 4%. The curves represent the normal distributions that are fitted based on the prediction variation caused by jittering all the samples in each of the 30 model repetitions.

3.3 Model interpretation

Finally, we focus on interpreting the optimal model (i.e., the weight-optimized multi-task neural network), which allows us to better understand the nature of the relationship between the total REE content (i.e., output of the model) and the contents of the six selected major oxides (i.e., Al_2O_3, Fe_2O_3, CaO, P_2O_5, TiO_2, and MgO). To this end, we conduct SHAP analysis (see details in Sec. 2.6), which is designed to interpret the latent data patterns learned by a "black box" machine learning model (*75*, *76*). In brief, the SHAP analysis quantifies the marginal contribution of the individual input features to the prediction target (based on the actual samples in the dataset), so as to capture the importance of each input.

Figure 4a displays a summary of the SHAP analysis for each input oxide in predicting the total REE content (ranked in terms of descending influence from top to bottom). Here, a positive marginal

contribution corresponds to an increase in the total REE content, and vice versa. The horizontal dispersion associated with each input in Fig. 4a captures the range of the marginal contribution values, their vertical width reflects the distribution of each feature in the coal ash dataset, and their colors indicate the normalized value of each input feature (increasing from blue to red). Based on this summary plot, we further extract the relative importance of each oxide in predicting the total REE content, as shown in Fig. 4b. In detail, the importance of each feature here is calculated by summing the marginal contribution of each oxide (over all the samples) and then normalizing the absolute value of the results so that the most important feature has an importance of 1. Overall, we find that Al_2O_3 is the most important feature and exhibits a positive influence on the total REE content. In contrast, the SHAP analysis highlights that CaO shows a negative influence on the total REE content. Although the other four oxides all exhibit some positive influences, they are notably less influential than the first two oxides.

To further interpret our multi-task neural network, we track the one-dimensional effect of each of the input oxides on the predicted total REE content. To this end, one at a time, we jitter the value of each input feature within its 20th-to-80th percentile in the coal ash dataset while fixing the other features to their median values (see further details in Sec. 2.6). This allows us to evaluate how the variation of each feature alters the model prediction. Based on the 30 repetitions, the results of the feature effect analysis are collectively displayed in Fig. 4c.

Here, we first observe that Al_2O_3 shows a strong quasi-linear positive contribution to the total REE content. This observation echoes the correlation analyses (see Figs. 1a and S2a) and the SHAP analysis (see Fig. 4a). This linear mapping also partially explains why the linear regression model considered herein (see Fig. 2c) is capable of achieving a decent prediction accuracy. In agreement with the SHAP analysis, a consistent negative correlation is observed between CaO and the total REE contents. Interestingly, we observe the existence of a discrepancy between the SHAP and feature effect analyses in the case of Fe_2O_3, wherein SHAP suggests a positive influence, but its feature effect curve has a negative slope. This may arise from the fact that, in contrast to SHAP, the outcome of the present one-dimensional feature effect analysis depends on the choice of the reference ash composition and does not capture potential coupling effects between the input features—so that Fe_2O_3 may exhibit either a positive or negative influence based on the specific ash that is considered.

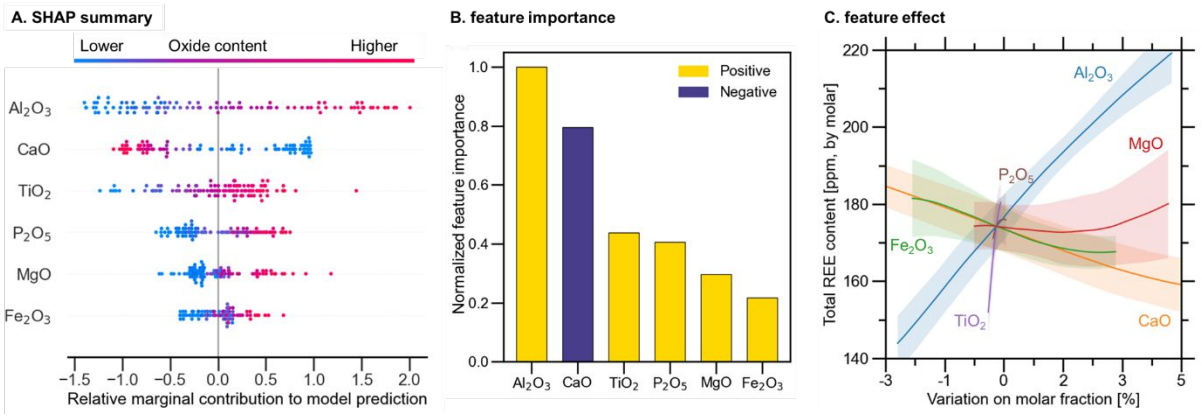

Figure 4: **Model interpretations. (A)** SHAP summary plot showing the marginal contribution of the input features (i.e., contents of the six major oxides) to the output (i.e., total REE content) for each sample in the coal ash dataset. **(B)** Normalized importance of the individual oxides based on the SHAP analysis. **(C)** Feature effect analysis, wherein each curve represents the mean prediction of the model after varying each oxide within its 20th-to-80th percentile in the coal ash dataset, and the shadowed region corresponds to the standard deviation (based on 30 repetitions of the model training).

4. Discussion

4.1 Improvement of the model prediction associated with the multi-task learning

We first discuss the origin of the boost in model performance associated with the multi-task approach. As compared to a conventional single-task paradigm that requires training 17 individual single-task models, our multi-task approach allows us to predict the 17 REE targets within a single model, which significantly reduces the time and effort for the model training. The results presented in Sec. 3.2 have important implications for both the effectiveness of the adopted dataset in training models and the improvement of model performance for high-throughput REE screening. Despite the limited size of the training data and the sole input of six oxides, our different models demonstrate comparably good testing accuracy, suggesting that the adopted dataset is sufficiently large for training decent models for high-throughput REE screening. The multi-task modeling approach directly increases the number of data points utilized during model training, thereby alleviating the overall level of scarcity. This increase in training data further contributes to enhanced model accuracy and suppressed noise sensitivity of the multi-task model. This can be attributed to the unique advantage of multi-task models, wherein sharing of artificial neurons in the hidden layers enables refining the noise sensitivity of a prediction task with the other prediction tasks that are less noise sensitive. Consequently, the multi-task model effectively achieves more reliable and accurate predictions.

Based on the correlation analysis (see Fig. 1a), we observe strong correlations among the contents of the individual REEs. This echoes the fact that different REE-bearing minerals are often embedded within the same glassy phase in coal ashes (*56*). This observation provides a theoretical basis for using a machine learning analysis to explore the prediction of REE contents—the individual REE contents lie within a well-defined low-dimensional manifold within the high-dimensional compositional space. (*81*). In comparison with knowledge-based approaches, machine learning models can exhibit superior performance in uncovering such low-dimensional folds and, therefore, can offer accurate predictions by only considering a small set of input features. This is generally a consequence of the fact that machine learning models excel at unveiling subtle patterns within datasets, that are otherwise invisible to humans. The above observation also indicates that the task of learning the mapping between the input primary oxide concentrations and the content of an individual REE can facilitate the task of predicting the content of another individual REE—which echoes the core concept of multi-task learning. In addition, we note that each REE comes with its noise, which can result in overfitting when a model is independently trained to predict the content of each REE. However, since the noise patterns associated with each REE are independent from each other, simultaneously predicting all the REE contents makes it possible for the multi-task model to filter out these individual noise contributions—since each REE output offers a baseline that presents the model from overfitting the noise associated with the other REEs.

Overall, the advantages of the multi-task approach can be summarized as follows (*57*): (i) **implicit data augmentation**: the multi-task model is trained with more samples, associated with diverse noise patterns; (ii) **attention focusing**: this approach helps the neural network to mitigate the influence associated with the limited size of the coal ash dataset and its high dimensionality of the input features; (iii) **eavesdropping**: some useful features for predicting the total REE content may be difficult to extract by the single-task model but easily uncovered from predicting the individual REE targets; (iv) **regularization**: the additional noise introduced by the 17 targets reduces the proclivity of overfitting for the neural network model, as some noise patterns are more detectable. These observations echo many material-related problems, wherein material properties often exhibit some partial level of correlation, but are associated with distinct noise (due to different measurements and test protocols, as well as randomness).

4.2 Correlation between bulk XRF oxides and the presence of REEs

The model interrogations reveal that data patterns learned by our model echo related knowledge established from previous research, which supports that the characteristic patterns learned by our model are indicative of coal ashes bearing elevated REE contents for the potential application of high-throughput screening. Recently, Chatterjee et al. also identified influential features to predict the total REE content in raw coals (*53*). Despite the differences in the target material (raw coal vs. coal ash here) and the selection of input features (here, we focus on the oxides that can be measured by XRF), we note that Al_2O_3 and Fe_2O_3 were also selected as being influential in that study. This suggests that those two oxides can indeed be influential in predicting the REE content in coal-related materials. Comparing those two oxides, Chatterjee et al. also observed that Al_2O_3 shows a larger influence than Fe_2O_3, which is in line with our findings from the SHAP analysis (see Fig. 4b).

Regarding feature-property mapping, our model reveals insightful patterns regarding the nature of the correlation between bulk XRF oxides and REEs. Previous studies suggested the existence of three sources of REEs in coal, namely, pyroclastic minerals, diagenetic minerals, and organics (*18*). Based on the model interpretation in Sec. 3.3, we find that Al_2O_3 features a positive influence on the total REE content in coal ashes. This observation echoes the findings from several previous studies suggesting that the presence of REEs is related to that of aluminous phases in coal ash (*13*, *28*, *38*, *39*), as well as in coal materials (*53*, *82*). Similarly, Querol et al. found that REEs are only present in the glassy aluminosilicate areas of the coal ash samples (*83*). Our model also indicates a positive correlation with P_2O_5, which is consistent with previous studies reporting that elevated concentrations of REEs may be linked to phosphates in coal ashes (*18*, *24*). In addition, the SHAP-based model interpretation also indicates a positive influence of Fe_2O_3 on the REE concentration. This is supported by previous observations showing that REEs tend to reside in Fe-rich aluminosilicates, Fe-rich multi-element eutectic, and Fe-oxides (*13–15*, *84*). Furthermore, we observe that CaO exhibits a strong negative correlation to the presence of REEs. This might be related to the origin of the coal, wherein lignite coals tend to contain more calcite than bituminous coals (*85*). In comparison, bituminous coal is a higher-rank coal with a higher carbon content, which is associated with the presence of REEs (*86*). Additionally, it is noteworthy that both CaO and Al_2O_3 are prominent components in coal ashes. A higher CaO content may indirectly imply a lower Al_2O_3 content, signifying a reduced presence of aluminous phases that typically contain REEs. The higher REE concentration in low-CaO coal ashes may take practical implications for recovering REEs from coal ashes—the remarkable concurrence between the low CaO and the high REE contents suggests that the CaO-poor legacy coal ashes (which have otherwise often been simply deposited in impoundments or landfills) may be a promising venue for REE extraction. This could facilitate the

614 recycling of out-of-specification coal ashes that cannot be used to replace cement, which, in turn, could
615 induce a paradigm shift in waste-to-resource management and circular economy.

616 *4.3 Potential for screening of high REE-bearing coal ashes*

617 In evaluating the applicability of our proposed approach for high-throughput screening of REE-
618 bearing coal ashes, we discuss the robustness of our multi-task neural network models in distinguishing
619 between ashes with low and high total REE contents. This distinction is crucial for the practical viability
620 of the screening application. While different studies have varying definitions of high REE-bearing ashes,
621 here we categorize the coal ash samples into two classes based on the median value of the total REE
622 content of the dataset, set at 157 ppm (see Table 1). As a demonstration, the effectiveness of the two multi-
623 task models presented in Sec. 3.2 in distinguishing these classes is evaluated using the intersection over
624 union (IoU) metric, and the results are summarized in Table 3.

625 In comparison, the IoU accuracy of the weight-optimized model surpasses that of its equal-weight
626 counterpart on both training and test sets. In terms of identifying high-REE samples in the test set, the
627 IoU increases from 64% to 70% when switching from the equal-weight model to the weight-optimized
628 model. Notably, the weight-optimized model demonstrates a precision of 77.7% and a recall of 87.5%,
629 which again underscores its capability in discriminating between low and high REE-bearing coal ashes.
630 Given the significance of precision for the efficiency of REE recovery (e.g., cost and time) and that of recall
631 for reliability capturing recovery opportunities, the relatively high precision and recall of this model hold
632 promise for real-world applications in screening high-REE-bearing coal ashes.

633

Table 3: **Confusion matrix and intersection over union (IoU) classification accuracy for the multi-task neural-network-based classification.** The threshold distinguishing low and high REE contents (i.e., 157 ppm) is determined based on the dataset's median total REE content.

Confusion matrix				True total REE content [ppm, by molar]		IoU accuracy [%]
				< 157	≥ 157	
Predictions by the multi-task neural net models [ppm, by molar]	Equal weight	Training set	< 157	39	3	83
			≥ 157	5	37	82
		Test set	< 157	4	1	50
			≥ 157	3	7	64
	Optimized weight	Training set	< 157	41	2	89
			≥ 157	3	38	88
		Test set	< 157	5	1	63
			≥ 157	2	7	70

4.4 Model transferability to other coal ash samples

While our proposed multi-task modeling approach exhibits good generalizability, a crucial concern for its real-world application in screening REE in coal ashes lies in its transferability and scalability across diverse origins, encompassing geological location, coal type, and processing protocols. Ideally, with sufficient data from a new operation, retraining the model from scratch would be straightforward. However, this is often unattainable due to the prohibitive cost and time required for measuring REE content. In that regard, a promising solution is implementing transfer learning (*87*). Given a sparse dataset, transfer learning involves retraining only a subset of the artificial neurons in the pretrained model, such that the transferred model can achieve desirable accuracy. For testing, we additionally curated a separate REE dataset using the raw data reported by Franus et al., which correspond to 12 coal ashes from 10 power plants in Poland (*88*). This separate dataset can be found in Supplementary Materials (Table S2).

Using the separate REE dataset, we compare the performance of three models. First, we assess our weight-optimized multi-task neural network (i.e., present model) based on all samples in this dataset. Second, we train a new multi-task neural network from scratch, while keeping the same architecture (i.e., non-transferred model). Third, we implement transfer learning by only further training the neurons in the output layer of the present model while keeping all others frozen (i.e., transferred model). For both

non-transferred and transferred models, we randomly selected 6 samples (50% of the separate dataset) for model training, following the pipeline detailed in Sec. 2.3. As a reference, all three models presented in Fig. 5 are accessible on GitHub (see Sec. S5).

As illustrated in Fig. 5a, our present model exhibits a consistent deviation in predicting the total REE content for the new samples, expected due to differences in the origins of the coal ash samples. Nevertheless, the highly linear distribution of the scatters in this plot indicates that this model still captures the overall trend of the total REE content of the separate dataset. The performance of the non-transferred and transferred models is compared in Figs. 5b and 5c, respectively. The non-transferred model essentially fails to generalize to the test samples—with only 6 samples available for training, preventing overfitting in the non-transferred model proved extremely challenging. In contrast, the transferred model exhibits good accuracy for samples in both the training and test sets, achieving an R^2 test accuracy of 0.86. Furthermore, we note that the transferred model corrects the systematic deviation associated with the present model (Fig. 5a). More importantly, the fact that the two clusters of samples between 100 and 200 ppm in Fig. 5a are all distributed along the line of equality in Fig. 5c, signifying that the transferred model effectively improves the predictions for individual samples, far surpassing the capabilities of a linear correction.

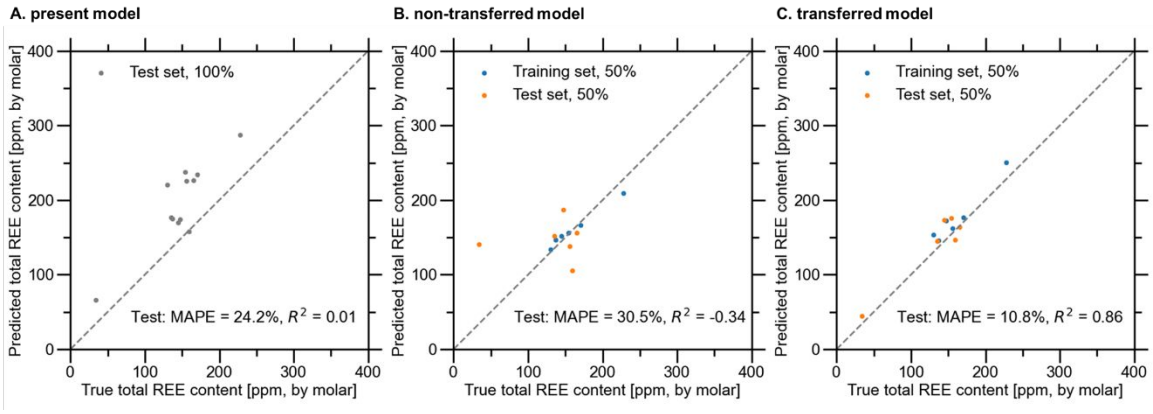

Figure 5: Transfer-learning-based refinement of the present model for a fully separated REE dataset (*88*)**.** Prediction accuracy of **(A)** the optimized multi-task neural network trained in this study (see Sec. 3.2), without any additional change; **(B)** the model with the same architecture but trained from scratch, with a random train-test split allocating half of the separate dataset for training; (**C)** the model with the same architecture but further trained using transfer learning, with the same train-test split. The $y = x$ dashed line indicates the perfect agreement.

As an additional validation of the transfer learning approach, we further delve into tuning the classification accuracy of the transferred model. Specifically, we penalize false positives and false negatives in predicting high-REE samples (≥ 157 ppm, see Sec. 4.3) in the separate dataset, wherein training and test set samples are not differentiated. To prove the concept, we adopt a straightforward approach by altering the loss function used for model training—by amplifying the original loss (with its squared value) when the predicted total REE content is either higher (i.e., penalize false positive) or lower (i.e., penalize false negative) than the true value. As shown in Table 4, we compare the classification performance between (i) the baseline transferred model shown in Fig. 5c and two additional transferred models trained penalizing (ii) false positive and (iii) false negative predictions. It can be seen that the second model reduces false positive predictions from 4 to 3, while the third model eliminates false negative predictions from 1 to 0.

These findings underscore the potential of transfer learning to enhance the adaptability of our multi-task modeling approach across diverse coal ash samples. Depending on the priority of classification emphasis, the transferred model can be finely tuned to enhance either the efficiency or reliability of REE recovery. Hence, our proposed approach provides a promising solution for coal-ash-based REE recovery.

Table 4: **Confusion matrix for three transferred models tuned for different targets in identifying high REE-bearing ashes in the separate dataset.** The threshold distinguishing low and high REE contents (i.e., 157 ppm) is kept the same as that from Table 3.

Confusion matrix			True total REE content [ppm, by molar]	
			< 157	≥ 157
Predictions by the three transferred models [ppm, by molar]	Baseline	< 157	4	1
		≥ 157	4	3
	Penalize false positive	< 157	5	3
		≥ 157	3	1
	Penalize false negative	< 157	1	0
		≥ 157	7	4

4.5 Further improvements of the modeling work

Despite the encouraging results obtained herein, our proposed model can be further enhanced in the following aspects. First of all, although the dataset is curated to represent the general compositional

variation of coal ashes within the U.S., it may not capture further varieties of ashes that are associated with other coal sources, types, and locations, which may exhibit systematic differences from those considered in this study (see Fig. 5a) (*18*). At this point, it is unclear whether coal ash samples from various locations (e.g., worldwide) can be modeled by a single model; however, it is clear that developing large, systematic, and reliable coal ash datasets is key for addressing the scarcity of REEs via high-throughput screening of coal ashes. It should be noted that the size of the dataset can also affect the level of confidence in our model. Previous studies have reported other uncertainties in predicting the REE content in coal ash (e.g., due to the coal origin, aging, interactions between individual REEs and different mineral phases, etc.) (*16*, *17*, *88*). Despite the efforts taken to enhance the data variety for training our models, the model may not always generalize outside the domain of the present dataset, even with the transfer learning approach discussed in Sec. 4.4. Therefore, increasing the size and scope of the dataset would be needed to further validate the developed model.

Improvements can be also implemented in terms of the choice of input features. Instead of using the oxide contents of coal ashes after calcination as model inputs, it would be more practical to use the oxide contents of raw coal ashes for large-scale screening in production. In that regard, our recent study reported that the coal ash content can be accurately predicted by combining XRF and machine learning methods (*89*). Such an approach could be considered to enable a direct prediction of REE concentrations based on the XRF measurement from the raw material, but would require additional data. In addition, we note that the REE contents may also be altered by other factors such as the type of coal combustion, unburnt residues, and processing condition of coal ashes. Hence, considering those non-chemical factors as input features is likely to further increase the robustness of the model.

Finally, to improve the transferability and scalability of the multi-task modeling to other material datasets, it is worth exploring algorithms for automating the determination of optimal weight distribution on different targets. In that regard, inspiration may be sought from the recent advances in the broad fields of computer vision and data informatics.

5. Conclusions

Overall, our study validates the feasibility of using a multi-task neural network model to fulfill accurate, real-time, and high-throughput screening of REE-rich coal ashes, based on the sole knowledge of the bulk composition. Our major findings are summarized below.

- The stratified sampling and recursive feature elimination techniques are demonstrated to be effective for enhancing the accuracy of machine learning models trained based on small and sparse datasets, as commonly encountered in materials science and engineering.
- The multi-task neural network models introduced herein exhibit consistently higher prediction accuracy and robustness than conventional single-task machine learning models. The multi-task models take advantage of the similarities between each of the individual input-output mappings to filter out noise associated with the individual REE measurement. Consequently, the prediction of the individual REEs synergically reinforces the prediction of the total REE content.
- Model interrogations reveal that Al_2O_3 and CaO contents have a positive and negative influence on the total REE content in coal ashes, respectively. This highlights the potential of extracting REEs from CaO-poor coal ashes, which are otherwise often simply deposited in impoundments due to their limited ability to replace cement in concrete.
- Transfer learning presents a promising solution for overcoming the often-limited data availability of REE measurement across different operations.
- Given the good transferability of machine learning approaches in material research, we envision that our proposed multi-task approach can be broadly applied to many other types of materials with sparse testing data ranging over multiple properties.

Acknowledgments

The authors acknowledge the financial support for this research by the U.S. Department of Transportation through the Federal Highway Administration (Grant #: 693JJ31950021) and the U.S. National Science Foundation (DMREF: 1922167).

Author contributions

B.G. was responsible for sample collections. Y.S., and Y.Z., were responsible for data processing and modeling analysis. Y.S., Y.Z., and A.G. drafted the manuscript. B.G., G.S., and M.B. reviewed the manuscript. B.G., G.S., and M.B. were responsible for the conception.

Declaration of interests

The authors declare no competing interests.

Data availability

The data used in this study can be shared upon reasonable request.

References

1. B. S. Van Gosen, P. L. Verplanck, R. R. Seal II, K. R. Long, J. Gambogi, "Rare-earth elements" (USGS Numbered Series 1802-O, U.S. Geological Survey, Reston, VA, 2017); https://doi.org/10.3133/pp1802O.

2. F. team, The role of rare earth elements in wind energy and electric mobility, *EU Science Hub - European Commission* (2020). https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/role-rare-earth-elements-wind-energy-and-electric-mobility.

3. E. Alonso, A. M. Sherman, T. J. Wallington, M. P. Everson, F. R. Field, R. Roth, R. E. Kirchain, Evaluating Rare Earth Element Availability: A Case with Revolutionary Demand from Clean Technologies. *Environ. Sci. Technol.* **46**, 3406–3414 (2012).

4. J. Rajesh Kumar, J.-Y. Lee, "Recovery of Critical Rare Earth Elements for Green Energy Technologies" in *Rare Metal Technology 2017*, H. Kim, S. Alam, N. R. Neelameggham, H. Oosterhof, T. Ouchi, X. Guan, Eds. (Springer International Publishing, Cham, 2017)*The Minerals, Metals & Materials Series*, pp. 19–29.

5. S. Hoenderdaal, L. Tercero Espinoza, F. Marscheider-Weidemann, W. Graus, Can a dysprosium shortage threaten green energy technologies? *Energy* **49**, 344–355 (2013).

6. "Understanding Generation and Storage Technology Supply Chain Risks and Needs to Support Electric Utility Sector Decarbonization" (Electric Power Research Institute, 2022); https://www.epri.com/research/products/000000003002023228.

7. A. Tukker, Rare Earth Elements Supply Restrictions: Market Failures, Not Scarcity, Hamper Their Current Use in High-Tech Applications. *Environ. Sci. Technol.* **48**, 9973–9974 (2014).

8. C. Hurst, "China's Rare Earth Elements Industry: What Can the West Learn?" (INSTITUTE FOR THE ANALYSIS OF GLOBAL SECURITY WASHINGTON DC, 2010); https://apps.dtic.mil/sti/citations/ADA525378.

9. C. Preinfalk, G. Morteani, "The Industrial Applications of Rare Earth Elements" in *Lanthanides, Tantalum and Niobium*, P. Möller, P. Černý, F. Saupé, Eds. (Springer, Berlin, Heidelberg, 1989)*Special Publication No. 7 of the Society for Geology Applied to Mineral Deposits*, pp. 359–370.

10. E. Gholz, "Rare Earth Elements and National Security" (Council on Foreign Relations, 2014); https://www.jstor.org/stable/resrep00311.

11. "Mineral commodity summaries 2020" (USGS Unnumbered Series, U.S. Geological Survey, Reston, VA, 2020); https://doi.org/10.3133/mcs2020.

12. FACT SHEET: Securing a Made in America Supply Chain for Critical Minerals, *The White House* (2022). https://www.whitehouse.gov/briefing-room/statements-releases/2022/02/22/fact-sheet-securing-a-made-in-america-supply-chain-for-critical-minerals/.

13. P. Liu, R. Huang, Y. Tang, Comprehensive Understandings of Rare Earth Element (REE) Speciation in Coal Fly Ashes and Implication for REE Extractability. *Environ. Sci. Technol.* **53**, 5369–5377 (2019).

14. W. Zhang, A. Noble, X. Yang, R. Honaker, A Comprehensive Review of Rare Earth Elements Recovery from Coal-Related Materials. *Minerals* **10**, 451 (2020).

15. Y. Sun, G. Qi, X. Lei, H. Xu, L. Li, C. Yuan, Y. Wang, Distribution and mode of occurrence of uranium in bottom ash derived from high-germanium coals. *Journal of Environmental Sciences* **43**, 91–98 (2016).

16. C. Scott, A. Kolker, "Rare earth elements in coal and coal fly ash" (2019–3048, U.S. Geological Survey, 2019); https://doi.org/10.3133/fs20193048.

17. R. S. Blissett, N. Smalley, N. A. Rowson, An investigation into six coal fly ashes from the United Kingdom and Poland to evaluate rare earth element content. *Fuel* **119**, 236–239 (2014).

18. V. V. Seredin, S. Dai, Coal deposits as potential alternative sources for lanthanides and yttrium. *International Journal of Coal Geology* **94**, 67–93 (2012).

19. DOE Launches $140 Million Program to Develop America's First-of-a-Kind Critical Minerals Refinery, *Energy.gov*. https://www.energy.gov/articles/doe-launches-140-million-program-develop-americas-first-kind-critical-minerals-refinery.

20. L. Fedele, J. A. Plant, B. D. Vivo, A. Lima, The rare earth element distribution over Europe: geogenic and anthropogenic sources. *Geochemistry: Exploration, Environment, Analysis* **8**, 3–18 (2008).

21. S. Dai, P. Xie, S. Jia, C. R. Ward, J. C. Hower, X. Yan, D. French, Enrichment of U-Re-V-Cr-Se and rare earth elements in the Late Permian coals of the Moxinpo Coalfield, Chongqing, China: Genetic implications from geochemical and mineralogical data. *Ore Geology Reviews* **80**, 1–17 (2017).

22. S. Park, M. Kim, Y. Lim, J. Yu, S. Chen, S. W. Woo, S. Yoon, S. Bae, H. S. Kim, Characterization of rare earth elements present in coal ash by sequential extraction. *Journal of Hazardous Materials* **402**, 123760 (2021).

23. A. Kumari, R. Parween, S. Chakravarty, K. Parmar, D. D. Pathak, J. Lee, M. K. Jha, Novel approach to recover rare earth metals (REMs) from Indian coal bottom ash. *Hydrometallurgy* **187**, 1–7 (2019).

24. Full article: A Review of the Occurrence and Promising Recovery Methods of Rare Earth Elements from Coal and Coal By-Products.

https://www.tandfonline.com/doi/full/10.1080/19392699.2015.1033097?casa_token=_wt-UqcrYJcAAAAA%3A-rnq_qmFbs3e0ST4rMFFDIOzvDky7-48GUf0iiLYNPtOxqot0ZUN5oAXwdDgfroE5c-QlSPzK34gDGo.

25. Full article: Sample Preparation for Determination of Rare Earth Elements in Geological Samples by ICP-MS: A Critical Review. https://www.tandfonline.com/doi/full/10.1080/00032719.2012.677778?casa_token=JCQ1C8g8u9IAAAAA%3AV2gh-VFJo--CytVIuM8hY5K0-j5vfnXe1Yq1Nqqd-YFUwIsmhJ-RbSgcmDuHv1SIe4n-ZQXKJu2bD2M.

26. X. Yan, S. Dai, I. T. Graham, X. He, K. Shan, X. Liu, Determination of Eu concentrations in coal, fly ash and sedimentary rocks using a cation exchange resin and inductively coupled plasma mass spectrometry (ICP-MS). *International Journal of Coal Geology* **191**, 152–156 (2018).

27. M. V. Balarama Krishna, G. Venkateswarlu, D. Karunasagar, Development of a simple and robust microwave-assisted decomposition method for the determination of rare earth elements in coal fly ash by ICP-OES. *Analytical Methods* **9**, 2031–2040 (2017).

28. A. Kolker, C. Scott, J. C. Hower, J. A. Vazquez, C. L. Lopano, S. Dai, Distribution of rare earth elements in coal combustion fly ash, determined by SHRIMP-RG ion microprobe. *International Journal of Coal Geology* **184**, 1–10 (2017).

29. T. X. Phuoc, P. Wang, D. McIntyre, Detection of rare earth elements in Powder River Basin sub-bituminous coal ash using laser-induced breakdown spectroscopy (LIBS). *Fuel* **163**, 129–132 (2016).

30. *232.2R-18: Report on the Use of Fly Ash in Concrete* (The American Concrete Institute, 2018).

31. T. Oey, J. Timmons, P. Stutzman, J. W. Bullard, M. Balonis, M. Bauchy, G. Sant, An improved basis for characterizing the suitability of fly ash as a cement replacement agent. *Journal of the American Ceramic Society* **100**, 4785–4800 (2017).

32. G. Xu, X. Shi, Characteristics and applications of fly ash as a sustainable construction material: A state-of-the-art review. *Resources, Conservation and Recycling* **136**, 95–109 (2018).

33. ASTM C618, "Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use in Concrete" (ASTM International, 2019); https://doi.org/10.1520/C0618-19.

34. J. M. Paris, J. G. Roessler, C. C. Ferraro, H. D. DeFord, T. G. Townsend, A review of waste products utilized as supplements to Portland cement in concrete. *Journal of Cleaner Production* **121**, 1–18 (2016).

35. B. Prasad, K. Sangita, Heavy Metal Pollution Index of Ground Water of an Abandoned Open Cast Mine Filled with Fly Ash: a Case Study. *Mine Water Environ* **27**, 265–267 (2008).

36. P. K. Sarswat, M. Leake, L. Allen, M. L. Free, X. Hu, D. Kim, A. Noble, G. H. Luttrell, Efficient recovery of rare earth elements from coal based resources: a bioleaching approach. *Materials Today Chemistry* **16**, 100246 (2020).

37. American Coal Ash Association, "Coal Combustion Product (CCP) production & use survey report" (2018).

38. R. L. Thompson, T. Bank, S. Montross, E. Roth, B. Howard, C. Verba, E. Granite, Analysis of rare earth elements in coal fly ash using laser ablation inductively coupled plasma mass spectrometry and scanning electron microscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy* **143**, 1–11 (2018).

39. S. N. Montross, C. A. Verba, H. L. Chan, C. Lopano, Advanced characterization of rare earth element minerals in coal utilization byproducts using multimodal image analysis. *International Journal of Coal Geology* **195**, 362–372 (2018).

40. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).

41. K. M. Jablonka, D. Ongari, S. M. Moosavi, B. Smit, Big-Data Science in Porous Materials: Materials Genomics and Machine Learning. *Chem. Rev.*, acs.chemrev.0c00004 (2020).

42. K. Guo, Z. Yang, C.-H. Yu, M. J. Buehler, Artificial intelligence and machine learning in design of mechanical materials. *Mater. Horiz.*, doi: 10.1039/D0MH01451F (2020).

43. R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, Machine learning in materials informatics: recent applications and prospects. *npj Computational Materials* **3**, 1–13 (2017).

44. H. Liu, Z. Fu, K. Yang, X. Xu, M. Bauchy, Machine learning for glass science and engineering: A review. *Journal of Non-Crystalline Solids: X* **4**, 100036 (2019).

45. Y. Song, Z. Huang, C. Shen, H. Shi, D. A. Lange, Deep learning-based automated image segmentation for concrete petrographic analysis. *Cement and Concrete Research* **135**, 106118 (2020).

46. B. A. Young, A. Hall, L. Pilon, P. Gupta, G. Sant, Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods. *Cement and Concrete Research* **115**, 379–388 (2019).

47. N. M. Anoop Krishnan, S. Mangalathu, M. M. Smedskjaer, A. Tandia, H. Burton, M. Bauchy, Predicting the dissolution kinetics of silicate glasses using machine learning. *Journal of Non-Crystalline Solids* **487**, 37–45 (2018).

48. J. Schmidt, M. R. G. Marques, S. Botti, M. A. L. Marques, Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials* **5**, 1–36 (2019).

49. B. Ouyang, Y. Song, Y. Li, F. Wu, H. Yu, Y. Wang, G. Sant, M. Bauchy, Predicting Concrete's Strength by Machine Learning: Balance between Accuracy and Complexity of Algorithms. *MJ* **117**, 125–133 (2020).

50. A. G. Kusne, T. Gao, A. Mehta, L. Ke, M. C. Nguyen, K.-M. Ho, V. Antropov, C.-Z. Wang, M. J. Kramer, C. Long, I. Takeuchi, On-the-fly machine-learning for high-throughput experiments: search for rare-earth-free permanent magnets. *Scientific Reports* **4**, 6367 (2014).

51. L. Yang, C. Zhu, Y. Sheng, H. Nian, Q. Li, P. Song, W. Lu, J. Yang, B. Liu, Investigation of mechanical and thermal properties of rare earth pyrochlore oxides by first-principles calculations. *Journal of the American Ceramic Society* **102**, 2830–2840 (2019).

52. Y. Song, K. Yang, J. Chen, K. Wang, G. Sant, M. Bauchy, Machine Learning Enables Rapid Screening of Reactive Fly Ashes Based on Their Network Topology. *ACS Sustainable Chem. Eng.*, doi: 10.1021/acssuschemeng.0c06978 (2021).

53. S. Chatterjee, M. Mastalerz, A. Drobniak, C. Ö. Karacan, Machine learning and data augmentation approach for identification of rare earth element potential in Indiana Coals, USA. *International Journal of Coal Geology* **259**, 104054 (2022).

54. Y. Zhang, Q. Yang, An overview of multi-task learning. *National Science Review* **5**, 30–43 (2018).

55. H. Borchani, G. Varando, C. Bielza, P. Larrañaga, A survey on multi-output regression. *WIREs Data Mining and Knowledge Discovery* **5**, 216–233 (2015).

56. E. Moniz, "Report on Rare Earth Elements from Coal and Coal By-products" (U.S. Department of Energy, Washington, DC, 2017).

57. S. Ruder, An Overview of Multi-Task Learning in Deep Neural Networks. *arXiv:1706.05098 [cs, stat]* (2017).

58. T. L. Bank, E. A. Roth, P. Tinker, E. Granite, "Analysis of Rare Earth Elements in Geologic Samples using Inductively Coupled Plasma Mass Spectrometry; US DOE Topical Report - DOE/NETL-2016/1794" (NETL-PUB-20441, National Energy Technology Lab. (NETL), Pittsburgh, PA, (United States), 2016); https://doi.org/10.2172/1415779.

59. P. T. Durdziński, M. Ben Haha, S. A. Bernal, N. De Belie, E. Gruyaert, B. Lothenbach, E. Menéndez Méndez, J. L. Provis, A. Schöler, C. Stabler, Z. Tan, Y. Villagrán Zaccardi, A. Vollpracht, F. Winnefeld, M. Zając, K. L. Scrivener, Outcomes of the RILEM round robin on degree of reaction of slag and fly ash in blended cements. *Mater Struct* **50**, 135 (2017).

60. P. M. Granitto, C. Furlanello, F. Biasioli, F. Gasperi, Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemometrics and Intelligent Laboratory Systems* **83**, 83–90 (2006).

61. M. Gevrey, I. Dimopoulos, S. Lek, Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling* **160**, 249–264 (2003).

62. Y. P. Liu, M. G. Wu, J. X. Qian, Predicting coal ash fusion temperature based on its chemical composition using ACO-BP neural network. *Thermochimica Acta* **454**, 64–68 (2007).

63. W. Liang, G. Wang, X. Ning, J. Zhang, Y. Li, C. Jiang, N. Zhang, Application of BP neural network to the prediction of coal ash melting characteristic temperature. *Fuel* **260**, 116324 (2020).

64. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems* **32** (2019).

65. T. Chen, C. Guestrin, "XGBoost: A Scalable Tree Boosting System" in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016; http://arxiv.org/abs/1603.02754), pp. 785–794.

66. S. Santurkar, D. Tsipras, A. Ilyas, A. Madry, How Does Batch Normalization Help Optimization? *arXiv:1805.11604 [cs, stat]* (2019).

67. J. Howard, S. Gugger, *Deep Learning for Coders with Fastai and PyTorch* (O'Reilly Media, Inc., 2020).

68. J. Heaton, *Introduction to Neural Networks with Java* (Heaton Research, Inc., 2008).

69. T. Evgeniou, M. Pontil, "Regularized multi--task learning" in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, New York, NY, USA, 2004; https://doi.org/10.1145/1014052.1014067)*KDD '04*, pp. 109–117.

70. D. Xu, Y. Shi, I. W. Tsang, Y.-S. Ong, C. Gong, X. Shen, Survey on Multi-Output Learning. *IEEE Transactions on Neural Networks and Learning Systems* **31**, 2409–2429 (2020).

71. T. Standley, A. Zamir, D. Chen, L. Guibas, J. Malik, S. Savarese, "Which Tasks Should Be Learned Together in Multi-task Learning?" in *International Conference on Machine Learning* (PMLR, 2020; http://proceedings.mlr.press/v119/standley20a.html), pp. 9120–9132.

72. C. Kuenneth, A. C. Rajan, H. Tran, L. Chen, C. Kim, R. Ramprasad, Polymer informatics with multi-task learning. *Patterns* **2**, 100238 (2021).

73. M. Bettinelli, P. Taina, Rapid analysis of coal fly ash by x-ray fluorescence spectrometry. *X-Ray Spectrometry* **19**, 227–232 (1990).

74. R. Guo, L. Zhang, Y. Zheng, H. Yao, Accurate and stable measurement of ash in coal by X-ray fluorescence spectrometry based on partial least squares. *Frontiers in Physics* **10** (2022).

75. S. M. Lundberg, S.-I. Lee, "A Unified Approach to Interpreting Model Predictions" in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett, Eds. (Curran Associates, Inc., 2017; http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf), pp. 4765–4774.

76. L. S. Shapley, A. E. Roth, *The Shapley Value: Essays in Honor of Lloyd S. Shapley* (Cambridge University Press, 1988).

77. Y. Song, Y. Wang, K. Wang, G. Sant, M. Bauchy, Decoding the genome of cement by Gaussian Process Regression. 7.

78. Materials informatics: From the atomic-level to the continuum - ScienceDirect. https://www.sciencedirect.com/science/article/pii/S1359645419300667?casa_token=XdTyH2s5rGkAAAAA:gAKBFu6NFuVtrlHmqYQvtdKz-C27S-7MQjRdBuBDlelmM6ggms45WytJaNhlF3PHhXLtcmvD8tmT.

79. B. Ouyang, Y. Song, Y. Li, F. Wu, H. Yu, Y. Wang, Z. Yin, X. Luo, G. Sant, M. Bauchy, Using machine learning to predict concrete's strength: learning from small datasets. *Eng. Res. Express* **3**, 015022 (2021).

80. Y. Zhou, F.-J. Chang, L.-C. Chang, I.-F. Kao, Y.-S. Wang, Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. *Journal of Cleaner Production* **209**, 134–145 (2019).

81. L. Melas-Kyriazi, "The Mathematical Foundations of Manifold Learning" (arXiv:2011.01307, arXiv, 2020); https://doi.org/10.48550/arXiv.2011.01307.

82. Z. Yang, Y. Li, Q. Lou, D. Liu, Y. Ning, S. Yang, Y. Tang, Y. Zhang, Z. Tang, X. Wang, Release of uranium and other trace elements from coal ash by (NH4)2SO4 activation of amorphous phase. *Fuel* **239**, 774–785 (2019).

83. X. Querol, J. Fernández-Turiel, A. López-Soler, Trace elements in coal and their behaviour during combustion in a large power station. *Fuel* **74**, 331–343 (1995).

84. Y. Sun, M. Wu, L. Zheng, B. Wang, Y. Wang, Uranium speciation in coal bottom ash investigated via X-ray absorption fine structure and X-ray photoelectron spectra. *Journal of Environmental Sciences* **74**, 88–94 (2018).

85. P. Catalfamo, S. Di Pasquale, F. Corigliano, L. Mavilia, Influence of the calcium content on the coal fly ash features in some innovative applications. *Resources, Conservation and Recycling* **20**, 119–125 (1997).

86. B. Fu, J. C. Hower, W. Zhang, G. Luo, H. Hu, H. Yao, A review of rare earth elements and yttrium in coal ash: Content, modes of occurrences, combustion behavior, and extraction methods. *Progress in Energy and Combustion Science* **88**, 100954 (2022).

87. Z. Liu, C. T. Wu, M. Koishi, Transfer learning of deep material network for seamless structure–property predictions. *Comput Mech* **64**, 451–465 (2019).

88. W. Franus, M. M. Wiatros-Motyka, M. Wdowin, Coal fly ash as a resource for rare earth elements. *Environ Sci Pollut Res* **22**, 9464–9474 (2015).

89. Z. Wen, H. Liu, M. Zhou, C. Liu, C. Zhou, Explainable machine learning rapid approach to evaluate coal ash content based on X-ray fluorescence. *Fuel* **332**, 125991 (2023).