



Materials Horizons

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

Journal:	<i>Materials Horizons</i>
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

SCHOLARONE™
Manuscripts

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.

1 *Full title*

2 **Predicting Rare Earth Elements Concentration in Coal Ashes with** 3 **Multi-Task Neural Networks**

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

5 ¹ Affiliation 1:

6 Physics of Amorphous and Inorganic Solids Laboratory (PARISlab)

7 5731B Boelter Hall, Department of Civil and Environmental Engineering, University of California Los Angeles,
8 CA 90095, USA.

9 ² Affiliation 2:

10 Laboratory for the Chemistry of Construction Materials (LC²)

11 5731J Boelter Hall, Department of Civil and Environmental Engineering, University of California Los Angeles,
12 CA 90095, USA.

13 ³ Affiliation 3:

14 Electric Power Research Institute (EPRI)

15 3420 Hillview Avenue, Palo Alto, CA 94304, USA.

16 ⁴ Affiliation 4:

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

18 ⁵ Affiliation 5:

19 Department of Materials Science and Engineering, University of California, Los Angeles, CA, USA.

20 ⁶ Affiliation 6:

21 California Nanosystems Institute, University of California, Los Angeles, CA, USA.

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

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

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

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

38

39

40

41 **1. Introduction**

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

56 Given the low supply and high demand for REEs from ore deposits in the predictable future, many
57 recent studies have paid special attention to identifying new sources of REEs. Among the various
58 available sources, coal ash (the burning residue of coal from power plants) has been considered as an
59 important potential source of REEs (13–18). In that regard, the Department of Energy (DOE) has allocated
60 \$140 million to recover REEs from coal ashes (19). The large global reserves of coal have attracted
61 numerous research efforts to extract REEs from coal ashes, whose worldwide average REE concentrations
62 are estimated at around 400 ppm (by mass) (18). In addition, recovering REEs from coal ashes has several
63 potential advantages over the ore-based extraction (14), which include obtaining higher concentrations
64 of heavy and critical REEs (18) (which are much lower in supply, higher in price, and projected for a
65 greater demand), incurring negligible mining cost (since coal ash is a byproduct from the coal
66 production), and precluding the need of handling radionuclide hazards from the ore deposits. In recent
67 years, a large number of studies have been published to promote the concept of extracting REEs from
68 coal ash, and the relevant topics broadly revolve around (i) geological distributions, depositional
69 settings, and resource characteristics of coal deposits containing REEs (17, 18, 20, 21), (ii) novel process

70 techniques that separate REEs from coal ashes (22–24), and (iii) advanced methods detecting the
71 concentration of REEs in coal ashes, such as inductively coupled plasma mass spectrometry (ICP-MS)
72 (25, 26), inductively coupled plasma optical emission spectrometry (ICP-OES) (27), SHRIMP-RG ion
73 microprobe (28), and laser-induced breakdown spectroscopy (LIBS) (29).

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

84 For both freshly produced coal ashes and those deposited in impoundments, a batch-wise screening
85 of the coal ashes is necessary for an effective REE extraction. This is because the REE concentration from
86 a given coal source is affected by the combustion protocols, processing techniques, as well as storage
87 conditions (13, 14, 37). Thus, to enable an efficient REE extraction, it is of special significance to develop
88 accurate, rapid, and high-throughput screening methodologies that can identify the coal ashes presenting
89 the highest potential for REE extraction, that is, the ashes featuring the largest REE content. From the
90 production's perspective, almost all the characterization techniques (e.g., ICP-MS or ICP-OES) for
91 determining the REE concentration in coal ash are time and cost-prohibitive for real-time ash screening.
92 In contrast, X-ray fluorescence (XRF) is a fast, convenient, and economical measurement that has been
93 widely adopted for testing the composition of coal ashes. However, XRF is only capable of measuring
94 the bulk contents of major elements (i.e., elements showing elevated concentration) and does not offer
95 the level of accuracy that is required to detect REEs. Nevertheless, previous studies have suggested the
96 existence of a certain correlation between the total REE content and the bulk composition of coal ashes
97 (28, 38, 39). For instance, REEs tend to concentrate more in the amorphous aluminosilicate phase in coal
98 ashes.

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

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

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

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

139 2. Experimental procedures

140 2.1 The coal ash dataset

141 The coal ash dataset adopted herein was sourced by the Electric Power Research Institute (EPRI), based
142 on the experimental work done at National Energy Technology Laboratory (NETL). This dataset comprises
143 99 entries, which include 18 coal samples and 81 coal ash samples. These samples were originally
144 investigated to represent the systematic variations in coal ash compositions across the United States of
145 America, which cover four coal ranks (i.e., bituminous, sub-bituminous, lignite, and blend), three types of
146 coal boilers (i.e., pulverized, cyclone, integrated coal gasification combined cycle), ten collection origins
147 (e.g., fly ash, bottom ash, pond ash, and landfill ash), and nine different coal basins. All the samples were
148 measured following the procedure described in DOE/NETL-2016/1794 report (58). In detail, for either the
149 raw coal or coal ash samples, the moisture and organic compounds in the sample were first removed by
150 calcination in a high-temperature furnace at 1100°C. After that, the burning residue was ground into fine
151 powders that are less than 74 μm (200 mesh) in diameter, fused with lithium metaborate, and further fully
152 dissolved in concentrated acid solutions. The concentration of each element in the sample was then
153 measured using inductively coupled plasma mass spectrometry (ICP-MS), and the results were reported in
154 terms of the mass fraction of the individual elements in the raw sample (i.e., either coal or coal ash). For
155 each sample, the raw dataset reports the mass contents for a total of 51 major and trace elements, the latter
156 of which include all the 16 REEs except for promethium (Pm). To filter out the effect of the molar mass of
157 each element from the raw ICP data, we convert all the mass contents of the 51 elements measured from
158 the calcinated samples into the molar contents of their corresponding oxides. The molar contents of those
159 oxides are then renormalized based on the total oxide quantity in each sample.

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

170 *2.2 Techniques for addressing the sparse nature of the coal ash dataset*

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

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

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

194 2.3 Neural networks modeling

195 We adopt artificial neural networks for predicting the output target (i.e., the total REE content) based
196 on the input features (i.e., the contents of common oxides that can be easily measured by XRF). Here, we
197 specifically adopt neural networks as a regression model because of their good flexibility in handling
198 different types of datasets and interpolating complex relationships between the input features and output
199 target (61). This approach has been previously adopted for studying many glass-based materials (41, 43,
200 44, 48), including coal ashes (62, 63). Herein, the modeling process follows one of our previous studies
201 that also used neural networks to analyze the chemical reactivity of coal ashes (52), wherein the basics of
202 neural networks were thoroughly reviewed. The following contents focus on the core concepts of neural
203 networks and the modeling details exclusive to the investigation of REE prediction. The modeling work,
204 along with the analyses, in this study is conducted in Spyder IDE (version 5.4.3) with Python (version
205 3.10.10), and the specific program libraries are further detailed in the text. All the artificial neural
206 networks presented herein are implemented using the PyTorch deep learning library (version 2.0.0) (64).
207 The other machine learning models investigated involved in this study are built using the scikit-learn
208 (version 1.3.0) and XGBoost (version 1.7.6) libraries, as further detailed in Sec. 3.2 (40, 65). The core code of
209 the modeling work is available on GitHub (see Sec. S5).

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

218 In detail, the artificial neurons in the first hidden layer are designed to extract low-level features
219 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).

250 We carry out the hyperparameter selection based on a grid search, by systematically varying the
251 three hyperparameters within their respective reasonable ranges. The number of artificial neurons is
252 varied from 2 to 32 (with an incremental factor of 2) in each layer; learning rate is varied from 10^{-5} to 10^0
253 (with an incremental factor of $10^{0.33}$); and weight decay is varied from 10^{-6} to 10^{-1} (with an incremental factor
254 of $10^{0.33}$). Following common practices in training machine learning models (43, 68), we split a small
255 portion (20%) from the training set to validate the model performance based on the selected
256 hyperparameters. Based on the model accuracy observed from the validation samples, we determine the
257 optimal hyperparameters as (i) 20 and 6 neurons in the two hidden layers, (2) $10^{-3.33}$ for learning rate, and
258 10^{-2} for the weight decay. Based on additional model comparisons, further increasing the model complexity,
259 such as using more hidden layers, artificial neurons, or additional functional layers, does not result in
260 notable improvements in the model performance. As a demonstration, the variation of model accuracy with
261 different hidden layers is shown in Supplementary Materials (Sec. S2).

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

269 *2.4 Multi-task neural networks*

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

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

292 *2.5 Evaluation of noise sensitivity of the single and multi-task neural networks*

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

303 *2.6 Interpretation of the trained machine learning model*

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

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

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

321 **3. Results**

322 *3.1 Features selection*

323 To investigate the REE concentration in coal ashes, we curate a dataset comprising representative
324 99 coal ash samples, as detailed in Sec. 2.1. We select 10 major oxides to represent the bulk composition
325 of each sample (i.e., model inputs), while the target of the model prediction is the molar contents of 16
326 REEs (i.e., model outputs; note that Pm is not considered as this element is practically non-existent in
327 nature). We further use their sum as the ultimate prediction target (i.e., total REE content). The statistics of
328 this dataset are summarized in Table 1, and the correlation between the 16 individual REEs and the total
329 REE content is displayed in Fig. 1a. Here, we observe some strong correlations between the contents of
330 different REEs, which agree with the general idea that different types of REEs tend to be found within
331 same chemical phases in coal ashes (56). More importantly, several experimental studies observed certain
332 correlations between the presence of REEs in coal ashes and the bulk composition of some major elements
333 (29, 39, 40). These observations provide a ground for mapping the bulk chemical composition of coal
334 ashes to REE contents with machine learning, which is known to excel in capturing implicit data patterns
335 without explicit knowledge (41).

336

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

Statistics		Oxide fraction [molar%]											Total REE [ppm, by molar]
		SiO ₂	Al ₂ O ₃	CaO	Fe ₂ O ₃	MgO	Na ₂ O	K ₂ O	TiO ₂	P ₂ O ₅	MnO	Sum	
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

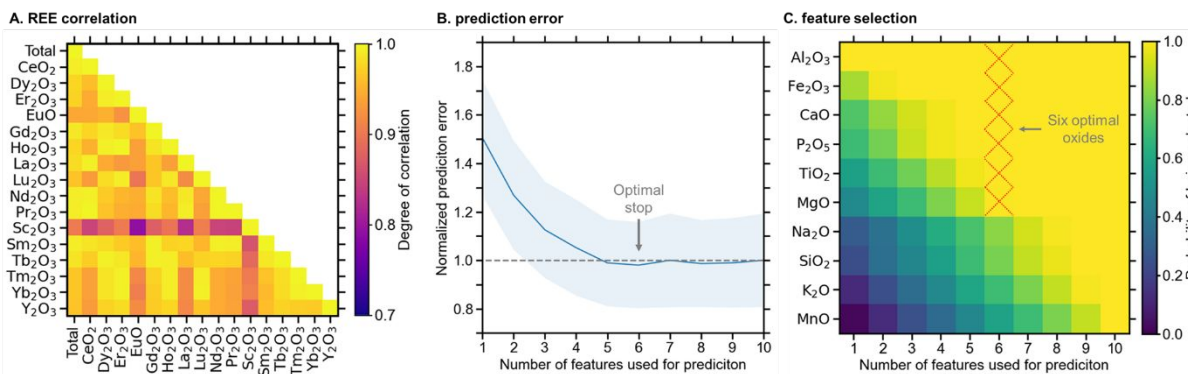
340 A major challenge of the machine learning analysis conducted in this study arises from the fact that
 341 the curated coal ash dataset has a relatively high dimensionality of input features (i.e., 10 by default; see
 342 Table 1) as compared to its small number of datapoints (99 samples), thereby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

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

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

363 The results of the RFE analysis are given in Figs. 1b and 1c. To ease the discussion, the prediction error
364 displayed in Fig. 1b is normalized based on the mean absolute percentage error (MAPE) value achieved
365 when all the ten features are used for training the model (i.e., when the model is trained with the maximum
366 amount of information). With the features being iteratively added to the model, we initially observe a
367 decline in the prediction error (and, hence, an improvement of the model accuracy) up to using six features.
368 However, the error then flattens out when more than six features are used in the model. Eventually, we
369 observe a slight increase in the prediction error. This suggests that these additional features do not provide
370 meaningful information for improving the model accuracy and, rather, unnecessarily increase the risk of
371 overfitting by increasing the model complexity and level of noise that the model is exposed to. By repeating
372 the RFE analysis 100 times, we then determine the probability for each input feature (i.e., each oxide) to be
373 selected as a function of the number of features being considered, that is, as a function of the RFE iteration
374 step (see Fig. 1c). We observe that the oxides that are selected early on (i.e., even when the number of
375 considered inputs is small) tend to continue to be selected after adding more features. This suggests that
376 this feature selection analysis yields a stable selection of the oxides that are the most informative to the
377 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
378 be more influential than the remaining four ones to infer the presence (or absence) of the REE-bearing
379 phases. However, the results in Fig. 1c should not be interpreted in a way such that the unselected features
380 are insignificant—rather, they are simply not contributing to further boost the prediction accuracy (e.g., due
381 to a high correlation with the selected features).

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



388

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

397 3.2 Single-task vs. multi-task neural network models

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

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

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

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

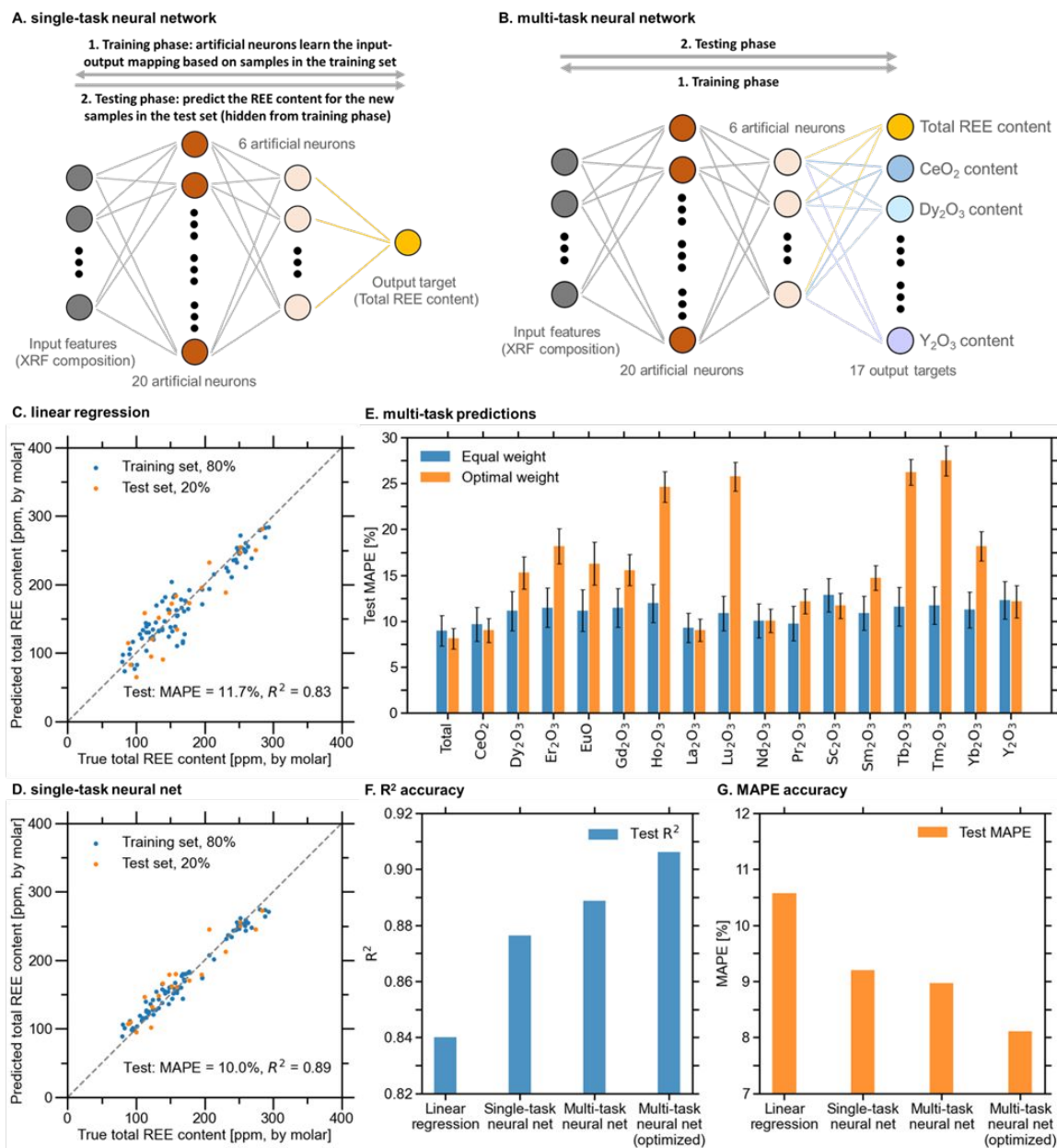
Stratification		Multivariate linear regression		Single-task neural net	
		No	Yes	No	Yes
Training set	R^2	0.880 ± 0.014	0.876 ± 0.010	0.904 ± 0.041	0.940 ± 0.019
	MAPE [%]	9.6 ± 0.1	9.8 ± 0.5	8.5 ± 1.7	6.7 ± 1.1
Test set	R^2	0.804 ± 0.087	0.847 ± 0.060	0.819 ± 0.119	0.875 ± 0.037
	MAPE [%]	11.1 ± 2.1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

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

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

439 as achieved by the two multi-task models, based on the same train-test splits as those used for the single-
440 task neural network. As a key result, we find that the two multi-task models both exhibit a notably
441 enhanced accuracy as compared to that achieved by the single-task model (see the accuracy comparison
442 between all the models in Figs. 2f and 2g). In detail, we then note that the model with equal weights tends
443 to accurately predict all the targets, with an average MAPE of approximately 10%. In particular, it
444 achieves a higher test accuracy and a lower standard deviation in predicting the total REE content than
445 the single-task models (see Table 2), where the average MAPE decreases from 9.5% to 9.0%. Remarkably,
446 the prediction accuracy scores shown in Fig. 2e are achieved by a single model, whereas predicting these
447 17 outputs would have required 17 distinct neural networks if using conventional single-task models.
448 Importantly, even when using the same weight for all the input features, the multi-task approach
449 outperforms the single-task approach in predicting the total REE content. This broadly highlights the
450 advantages of the multi-task approach (57), as further discussed in Sec. 4.1.

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



463

464

465

466

467

468

469

470

471

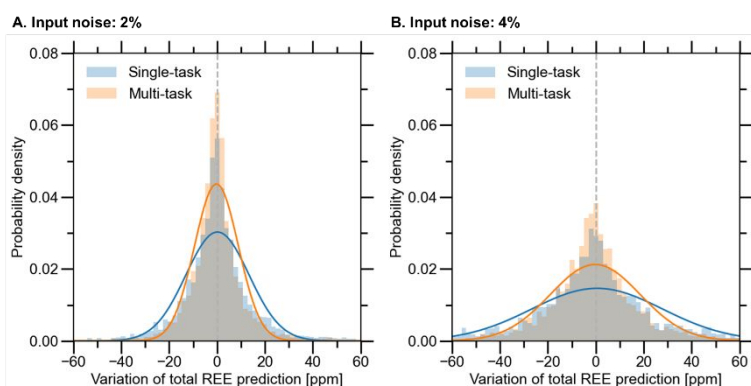
472

473

474

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.

475 The results above suggest that, by leveraging the information from multiple prediction targets, the
 476 multi-task model exhibits not only higher accuracy but also less variation than the single-task model in
 477 predicting the total REE content. A follow-up investigation is carried out to assess the sensitivity of the
 478 multi-task model to the noise from the model input. This is pertinent to the high-throughput screening
 479 of coal ashes, as XRF measurements are usually subject to variations (e.g. relative standard deviations of
 480 1 to 5% for major elements in coal ashes) (73, 74). Here, we introduce 2% and 4% Gaussian noise into each
 481 of the model inputs (see details in Sec. 2.5). The variations of the total REE prediction are shown in Fig.
 482 3. While the variations of the model prediction are normally distributed, the multi-task model yields
 483 notably higher peaks and narrower spreads of the prediction variations at both the noise levels. These
 484 results demonstrate that the multi-task model does exhibit improved resistance to the input noise.
 485 Further discussion about this improvement is provided in Sec. 4.1.



486

487 **Figure 3: Comparison of sensitivity to input noise between single- and multi-task neural networks**
 488 **for predicting the total REE content under different levels of noise.** Input noise levels are set at (A)
 489 2%, and (B) 4%. The curves represent the normal distributions that are fitted based on the prediction
 490 variation caused by jittering all the samples in each of the 30 model repetitions.

491 3.3 Model interpretation

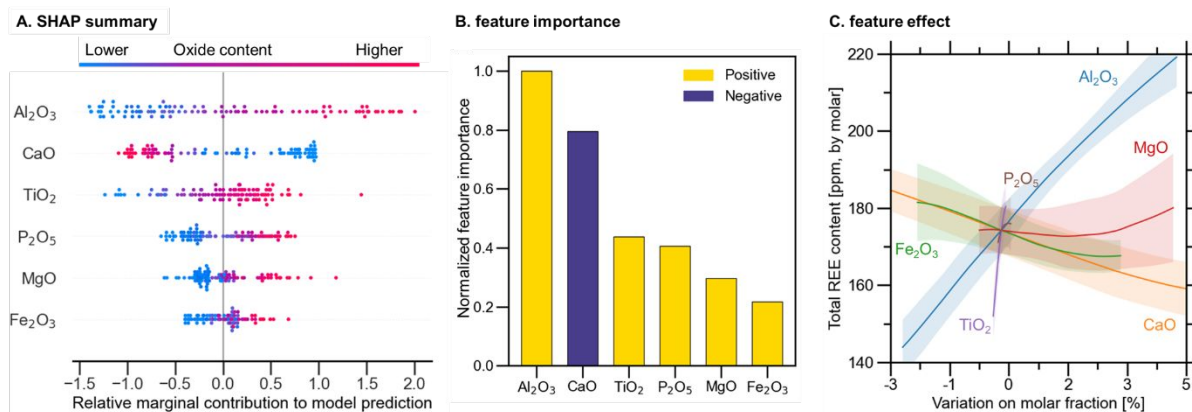
492 Finally, we focus on interpreting the optimal model (i.e., the weight-optimized multi-task neural
 493 network), which allows us to better understand the nature of the relationship between the total REE
 494 content (i.e., output of the model) and the contents of the six selected major oxides (i.e.,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

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

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

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

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



529

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

536 4. Discussion

537 4.1 Improvement of the model prediction associated with the multi-task learning

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

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

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

581 *4.2 Correlation between bulk XRF oxides and the presence of REEs*

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

592 Regarding feature-property mapping, our model reveals insightful patterns regarding the nature of
593 the correlation between bulk XRF oxides and REEs. Previous studies suggested the existence of three
594 sources of REEs in coal, namely, pyroclastic minerals, diagenetic minerals, and organics (18). Based on
595 the model interpretation in Sec. 3.3, we find that Al_2O_3 features a positive influence on the total REE
596 content in coal ashes. This observation echoes the findings from several previous studies suggesting that
597 the presence of REEs is related to that of aluminous phases in coal ash (13, 28, 38, 39), as well as in coal
598 materials (53, 82). Similarly, Querol et al. found that REEs are only present in the glassy aluminosilicate
599 areas of the coal ash samples (83). Our model also indicates a positive correlation with P_2O_5 , which is
600 consistent with previous studies reporting that elevated concentrations of REEs may be linked to
601 phosphates in coal ashes (18, 24). In addition, the SHAP-based model interpretation also indicates a
602 positive influence of Fe_2O_3 on the REE concentration. This is supported by previous observations
603 showing that REEs tend to reside in Fe-rich aluminosilicates, Fe-rich multi-element eutectic, and Fe-
604 oxides (13–15, 84). Furthermore, we observe that CaO exhibits a strong negative correlation to the
605 presence of REEs. This might be related to the origin of the coal, wherein lignite coals tend to contain
606 more calcite than bituminous coals (85). In comparison, bituminous coal is a higher-rank coal with a
607 higher carbon content, which is associated with the presence of REEs (86). Additionally, it is noteworthy
608 that both CaO and Al_2O_3 are prominent components in coal ashes. A higher CaO content may indirectly
609 imply a lower Al_2O_3 content, signifying a reduced presence of aluminous phases that typically contain
610 REEs. The higher REE concentration in low-CaO coal ashes may take practical implications for recovering
611 REEs from coal ashes—the remarkable concurrence between the low CaO and the high REE contents
612 suggests that the CaO-poor legacy coal ashes (which have otherwise often been simply deposited in
613 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

634 **Table 3: Confusion matrix and intersection over union (IoU) classification accuracy for the multi-task**
 635 **neural-network-based classification.** The threshold distinguishing low and high REE contents (i.e., 157
 636 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

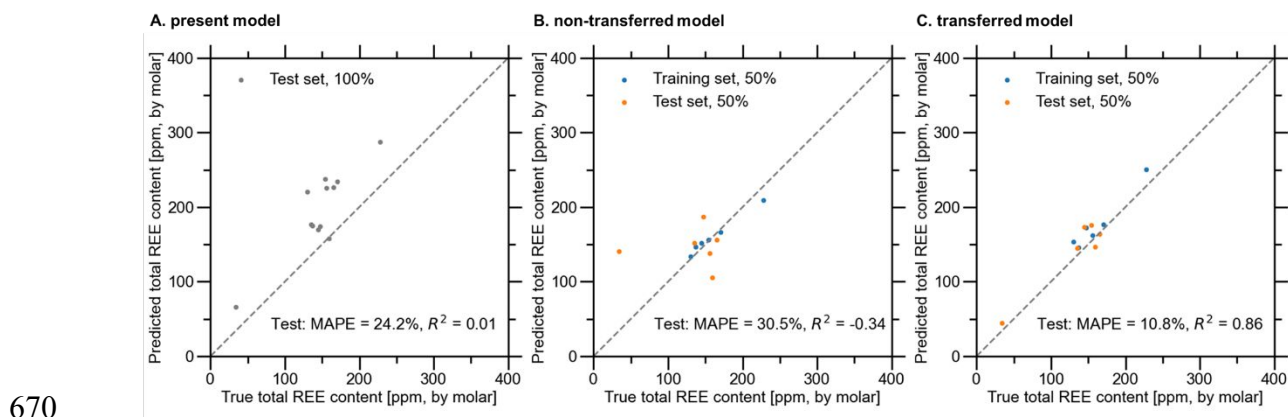
637 4.4 Model transferability to other coal ash samples

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

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

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

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



671 **Figure 5: Transfer-learning-based refinement of the present model for a fully separated REE dataset**
 672 (88). Prediction accuracy of (A) the optimized multi-task neural network trained in this study (see Sec.
 673 3.2), without any additional change; (B) the model with the same architecture but trained from scratch,
 674 with a random train-test split allocating half of the separate dataset for training; (C) the model with the
 675 same architecture but further trained using transfer learning, with the same train-test split. The $y = x$
 676 dashed line indicates the perfect agreement.

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

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

692 **Table 4: Confusion matrix for three transferred models tuned for different targets in identifying high**
 693 **REE-bearing ashes in the separate dataset.** The threshold distinguishing low and high REE contents (i.e.,
 694 157 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

695 4.5 Further improvements of the modeling work

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

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

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

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

723 5. Conclusions

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

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

744 **Acknowledgments**

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

748 **Author contributions**

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

752 **Declaration of interests**

753 The authors declare no competing interests.

754 **Data availability**

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

756

757 **References**

- 758 1. B. S. Van Gosen, P. L. Verplanck, R. R. Seal II, K. R. Long, J. Gambogi, "Rare-earth elements" (USGS
759 Numbered Series 1802-O, U.S. Geological Survey, Reston, VA, 2017);
760 <https://doi.org/10.3133/pp1802O>.
- 761 2. F. team, The role of rare earth elements in wind energy and electric mobility, *EU Science Hub -*
762 *European Commission* (2020). [https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-](https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/role-rare-earth-elements-wind-energy-and-electric-mobility)
763 [research-reports/role-rare-earth-elements-wind-energy-and-electric-mobility](https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/role-rare-earth-elements-wind-energy-and-electric-mobility).
- 764 3. E. Alonso, A. M. Sherman, T. J. Wallington, M. P. Everson, F. R. Field, R. Roth, R. E. Kirchain,
765 Evaluating Rare Earth Element Availability: A Case with Revolutionary Demand from Clean
766 Technologies. *Environ. Sci. Technol.* **46**, 3406–3414 (2012).
- 767 4. J. Rajesh Kumar, J.-Y. Lee, "Recovery of Critical Rare Earth Elements for Green Energy
768 Technologies" in *Rare Metal Technology 2017*, H. Kim, S. Alam, N. R. Neelameggham, H. Oosterhof,
769 T. Ouchi, X. Guan, Eds. (Springer International Publishing, Cham, 2017)*The Minerals, Metals &*
770 *Materials Series*, pp. 19–29.
- 771 5. S. Hoenderdaal, L. Tercero Espinoza, F. Marscheider-Weidemann, W. Graus, Can a dysprosium
772 shortage threaten green energy technologies? *Energy* **49**, 344–355 (2013).
- 773 6. "Understanding Generation and Storage Technology Supply Chain Risks and Needs to Support
774 Electric Utility Sector Decarbonization" (Electric Power Research Institute, 2022);
775 <https://www.epri.com/research/products/000000003002023228>.
- 776 7. A. Tukker, Rare Earth Elements Supply Restrictions: Market Failures, Not Scarcity, Hamper Their
777 Current Use in High-Tech Applications. *Environ. Sci. Technol.* **48**, 9973–9974 (2014).
- 778 8. C. Hurst, "China's Rare Earth Elements Industry: What Can the West Learn?" (INSTITUTE FOR
779 THE ANALYSIS OF GLOBAL SECURITY WASHINGTON DC, 2010);
780 <https://apps.dtic.mil/sti/citations/ADA525378>.
- 781 9. C. Preinfalk, G. Morteani, "The Industrial Applications of Rare Earth Elements" in *Lanthanides,*
782 *Tantalum and Niobium*, P. Möller, P. Černý, F. Saupé, Eds. (Springer, Berlin, Heidelberg, 1989)*Special*
783 *Publication No. 7 of the Society for Geology Applied to Mineral Deposits*, pp. 359–370.
- 784 10. E. Gholz, "Rare Earth Elements and National Security" (Council on Foreign Relations, 2014);
785 <https://www.jstor.org/stable/resrep00311>.
- 786 11. "Mineral commodity summaries 2020" (USGS Unnumbered Series, U.S. Geological Survey, Reston,
787 VA, 2020); <https://doi.org/10.3133/mcs2020>.

- 788 12. FACT SHEET: Securing a Made in America Supply Chain for Critical Minerals, *The White House*
789 (2022). [https://www.whitehouse.gov/briefing-room/statements-releases/2022/02/22/fact-sheet-](https://www.whitehouse.gov/briefing-room/statements-releases/2022/02/22/fact-sheet-securing-a-made-in-america-supply-chain-for-critical-minerals/)
790 [securing-a-made-in-america-supply-chain-for-critical-minerals/](https://www.whitehouse.gov/briefing-room/statements-releases/2022/02/22/fact-sheet-securing-a-made-in-america-supply-chain-for-critical-minerals/).
- 791 13. P. Liu, R. Huang, Y. Tang, Comprehensive Understandings of Rare Earth Element (REE) Speciation
792 in Coal Fly Ashes and Implication for REE Extractability. *Environ. Sci. Technol.* **53**, 5369–5377 (2019).
- 793 14. W. Zhang, A. Noble, X. Yang, R. Honaker, A Comprehensive Review of Rare Earth Elements
794 Recovery from Coal-Related Materials. *Minerals* **10**, 451 (2020).
- 795 15. Y. Sun, G. Qi, X. Lei, H. Xu, L. Li, C. Yuan, Y. Wang, Distribution and mode of occurrence of
796 uranium in bottom ash derived from high-germanium coals. *Journal of Environmental Sciences* **43**,
797 91–98 (2016).
- 798 16. C. Scott, A. Kolker, “Rare earth elements in coal and coal fly ash” (2019–3048, U.S. Geological
799 Survey, 2019); <https://doi.org/10.3133/fs20193048>.
- 800 17. R. S. Blissett, N. Smalley, N. A. Rowson, An investigation into six coal fly ashes from the United
801 Kingdom and Poland to evaluate rare earth element content. *Fuel* **119**, 236–239 (2014).
- 802 18. V. V. Seredin, S. Dai, Coal deposits as potential alternative sources for lanthanides and yttrium.
803 *International Journal of Coal Geology* **94**, 67–93 (2012).
- 804 19. DOE Launches \$140 Million Program to Develop America’s First-of-a-Kind Critical Minerals
805 Refinery, *Energy.gov*. [https://www.energy.gov/articles/doe-launches-140-million-program-](https://www.energy.gov/articles/doe-launches-140-million-program-develop-americas-first-kind-critical-minerals-refinery)
806 [develop-americas-first-kind-critical-minerals-refinery](https://www.energy.gov/articles/doe-launches-140-million-program-develop-americas-first-kind-critical-minerals-refinery).
- 807 20. L. Fedele, J. A. Plant, B. D. Vivo, A. Lima, The rare earth element distribution over Europe: geogenic
808 and anthropogenic sources. *Geochemistry: Exploration, Environment, Analysis* **8**, 3–18 (2008).
- 809 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
810 rare earth elements in the Late Permian coals of the Moxinpo Coalfield, Chongqing, China: Genetic
811 implications from geochemical and mineralogical data. *Ore Geology Reviews* **80**, 1–17 (2017).
- 812 22. S. Park, M. Kim, Y. Lim, J. Yu, S. Chen, S. W. Woo, S. Yoon, S. Bae, H. S. Kim, Characterization of
813 rare earth elements present in coal ash by sequential extraction. *Journal of Hazardous Materials* **402**,
814 123760 (2021).
- 815 23. A. Kumari, R. Parween, S. Chakravarty, K. Parmar, D. D. Pathak, J. Lee, M. K. Jha, Novel approach
816 to recover rare earth metals (REMs) from Indian coal bottom ash. *Hydrometallurgy* **187**, 1–7 (2019).
- 817 24. Full article: A Review of the Occurrence and Promising Recovery Methods of Rare Earth Elements
818 from Coal and Coal By-Products.

- 819 [https://www.tandfonline.com/doi/full/10.1080/19392699.2015.1033097?casa_token=_wt-](https://www.tandfonline.com/doi/full/10.1080/19392699.2015.1033097?casa_token=_wt-UqcrYJcAAAAA%3A-rnq_qmFbs3e0ST4rMFFDIOzvDky7-48GUf0iiLYNPtOxqot0ZUN5oAXwdDgfroE5c-QISPzK34gDGo)
820 [UqcrYJcAAAAA%3A-rnq_qmFbs3e0ST4rMFFDIOzvDky7-](https://www.tandfonline.com/doi/full/10.1080/19392699.2015.1033097?casa_token=_wt-UqcrYJcAAAAA%3A-rnq_qmFbs3e0ST4rMFFDIOzvDky7-48GUf0iiLYNPtOxqot0ZUN5oAXwdDgfroE5c-QISPzK34gDGo)
821 [48GUf0iiLYNPtOxqot0ZUN5oAXwdDgfroE5c-QISPzK34gDGo.](https://www.tandfonline.com/doi/full/10.1080/19392699.2015.1033097?casa_token=_wt-UqcrYJcAAAAA%3A-rnq_qmFbs3e0ST4rMFFDIOzvDky7-48GUf0iiLYNPtOxqot0ZUN5oAXwdDgfroE5c-QISPzK34gDGo)
- 822 25. Full article: Sample Preparation for Determination of Rare Earth Elements in Geological Samples
823 by ICP-MS: A Critical Review.
824 [https://www.tandfonline.com/doi/full/10.1080/00032719.2012.677778?casa_token=JCQ1C8g8u9IA](https://www.tandfonline.com/doi/full/10.1080/00032719.2012.677778?casa_token=JCQ1C8g8u9IAAAAA%3AV2gh-VFJo--CytVIuM8hY5K0-j5vfnXe1Yq1Nqqd-YFUwIsmhJ-RbSgcmDuHv1SIe4n-ZQXKJu2bD2M)
825 [AAAA%3AV2gh-VFJo--CytVIuM8hY5K0-j5vfnXe1Yq1Nqqd-YFUwIsmhJ-RbSgcmDuHv1SIe4n-](https://www.tandfonline.com/doi/full/10.1080/00032719.2012.677778?casa_token=JCQ1C8g8u9IAAAAA%3AV2gh-VFJo--CytVIuM8hY5K0-j5vfnXe1Yq1Nqqd-YFUwIsmhJ-RbSgcmDuHv1SIe4n-ZQXKJu2bD2M)
826 [ZQXKJu2bD2M.](https://www.tandfonline.com/doi/full/10.1080/00032719.2012.677778?casa_token=JCQ1C8g8u9IAAAAA%3AV2gh-VFJo--CytVIuM8hY5K0-j5vfnXe1Yq1Nqqd-YFUwIsmhJ-RbSgcmDuHv1SIe4n-ZQXKJu2bD2M)
- 827 26. X. Yan, S. Dai, I. T. Graham, X. He, K. Shan, X. Liu, Determination of Eu concentrations in coal, fly
828 ash and sedimentary rocks using a cation exchange resin and inductively coupled plasma mass
829 spectrometry (ICP-MS). *International Journal of Coal Geology* **191**, 152–156 (2018).
- 830 27. M. V. Balarama Krishna, G. Venkateswarlu, D. Karunasagar, Development of a simple and robust
831 microwave-assisted decomposition method for the determination of rare earth elements in coal fly
832 ash by ICP-OES. *Analytical Methods* **9**, 2031–2040 (2017).
- 833 28. A. Kolker, C. Scott, J. C. Hower, J. A. Vazquez, C. L. Lopano, S. Dai, Distribution of rare earth
834 elements in coal combustion fly ash, determined by SHRIMP-RG ion microprobe. *International*
835 *Journal of Coal Geology* **184**, 1–10 (2017).
- 836 29. T. X. Phuoc, P. Wang, D. McIntyre, Detection of rare earth elements in Powder River Basin sub-
837 bituminous coal ash using laser-induced breakdown spectroscopy (LIBS). *Fuel* **163**, 129–132 (2016).
- 838 30. *232.2R-18: Report on the Use of Fly Ash in Concrete* (The American Concrete Institute, 2018).
- 839 31. T. Oey, J. Timmons, P. Stutzman, J. W. Bullard, M. Balonis, M. Bauchy, G. Sant, An improved basis
840 for characterizing the suitability of fly ash as a cement replacement agent. *Journal of the American*
841 *Ceramic Society* **100**, 4785–4800 (2017).
- 842 32. G. Xu, X. Shi, Characteristics and applications of fly ash as a sustainable construction material: A
843 state-of-the-art review. *Resources, Conservation and Recycling* **136**, 95–109 (2018).
- 844 33. ASTM C618, “Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use in
845 Concrete” (ASTM International, 2019); <https://doi.org/10.1520/C0618-19>.
- 846 34. J. M. Paris, J. G. Roessler, C. C. Ferraro, H. D. DeFord, T. G. Townsend, A review of waste products
847 utilized as supplements to Portland cement in concrete. *Journal of Cleaner Production* **121**, 1–18
848 (2016).
- 849 35. B. Prasad, K. Sangita, Heavy Metal Pollution Index of Ground Water of an Abandoned Open Cast
850 Mine Filled with Fly Ash: a Case Study. *Mine Water Environ* **27**, 265–267 (2008).

- 851 36. P. K. Sarswat, M. Leake, L. Allen, M. L. Free, X. Hu, D. Kim, A. Noble, G. H. Luttrell, Efficient
852 recovery of rare earth elements from coal based resources: a bioleaching approach. *Materials Today*
853 *Chemistry* **16**, 100246 (2020).
- 854 37. American Coal Ash Association, "Coal Combustion Product (CCP) production & use survey
855 report" (2018).
- 856 38. R. L. Thompson, T. Bank, S. Montross, E. Roth, B. Howard, C. Verba, E. Granite, Analysis of rare
857 earth elements in coal fly ash using laser ablation inductively coupled plasma mass spectrometry
858 and scanning electron microscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy* **143**, 1–11 (2018).
- 859 39. S. N. Montross, C. A. Verba, H. L. Chan, C. Lopano, Advanced characterization of rare earth
860 element minerals in coal utilization byproducts using multimodal image analysis. *International*
861 *Journal of Coal Geology* **195**, 362–372 (2018).
- 862 40. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P.
863 Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, Scikit-learn: Machine
864 Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
- 865 41. K. M. Jablonka, D. Ongari, S. M. Moosavi, B. Smit, Big-Data Science in Porous Materials: Materials
866 Genomics and Machine Learning. *Chem. Rev.*, acs.chemrev.0c00004 (2020).
- 867 42. K. Guo, Z. Yang, C.-H. Yu, M. J. Buehler, Artificial intelligence and machine learning in design of
868 mechanical materials. *Mater. Horiz.*, doi: 10.1039/D0MH01451F (2020).
- 869 43. R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, Machine learning in
870 materials informatics: recent applications and prospects. *npj Computational Materials* **3**, 1–13 (2017).
- 871 44. H. Liu, Z. Fu, K. Yang, X. Xu, M. Bauchy, Machine learning for glass science and engineering: A
872 review. *Journal of Non-Crystalline Solids: X* **4**, 100036 (2019).
- 873 45. Y. Song, Z. Huang, C. Shen, H. Shi, D. A. Lange, Deep learning-based automated image
874 segmentation for concrete petrographic analysis. *Cement and Concrete Research* **135**, 106118 (2020).
- 875 46. B. A. Young, A. Hall, L. Pilon, P. Gupta, G. Sant, Can the compressive strength of concrete be
876 estimated from knowledge of the mixture proportions?: New insights from statistical analysis and
877 machine learning methods. *Cement and Concrete Research* **115**, 379–388 (2019).
- 878 47. N. M. Anoop Krishnan, S. Mangalathu, M. M. Smedskjaer, A. Tandia, H. Burton, M. Bauchy,
879 Predicting the dissolution kinetics of silicate glasses using machine learning. *Journal of Non-*
880 *Crystalline Solids* **487**, 37–45 (2018).

- 881 48. J. Schmidt, M. R. G. Marques, S. Botti, M. A. L. Marques, Recent advances and applications of
882 machine learning in solid-state materials science. *npj Computational Materials* **5**, 1–36 (2019).
- 883 49. B. Ouyang, Y. Song, Y. Li, F. Wu, H. Yu, Y. Wang, G. Sant, M. Bauchy, Predicting Concrete's
884 Strength by Machine Learning: Balance between Accuracy and Complexity of Algorithms. *MJ* **117**,
885 125–133 (2020).
- 886 50. A. G. Kusne, T. Gao, A. Mehta, L. Ke, M. C. Nguyen, K.-M. Ho, V. Antropov, C.-Z. Wang, M. J.
887 Kramer, C. Long, I. Takeuchi, On-the-fly machine-learning for high-throughput experiments:
888 search for rare-earth-free permanent magnets. *Scientific Reports* **4**, 6367 (2014).
- 889 51. L. Yang, C. Zhu, Y. Sheng, H. Nian, Q. Li, P. Song, W. Lu, J. Yang, B. Liu, Investigation of mechanical
890 and thermal properties of rare earth pyrochlore oxides by first-principles calculations. *Journal of the*
891 *American Ceramic Society* **102**, 2830–2840 (2019).
- 892 52. Y. Song, K. Yang, J. Chen, K. Wang, G. Sant, M. Bauchy, Machine Learning Enables Rapid Screening
893 of Reactive Fly Ashes Based on Their Network Topology. *ACS Sustainable Chem. Eng.*, doi:
894 10.1021/acssuschemeng.0c06978 (2021).
- 895 53. S. Chatterjee, M. Mastalerz, A. Drobniak, C. Ö. Karacan, Machine learning and data augmentation
896 approach for identification of rare earth element potential in Indiana Coals, USA. *International*
897 *Journal of Coal Geology* **259**, 104054 (2022).
- 898 54. Y. Zhang, Q. Yang, An overview of multi-task learning. *National Science Review* **5**, 30–43 (2018).
- 899 55. H. Borchani, G. Varando, C. Bielza, P. Larrañaga, A survey on multi-output regression. *WIREs Data*
900 *Mining and Knowledge Discovery* **5**, 216–233 (2015).
- 901 56. E. Moniz, "Report on Rare Earth Elements from Coal and Coal By-products" (U.S. Department of
902 Energy, Washington, DC, 2017).
- 903 57. S. Ruder, An Overview of Multi-Task Learning in Deep Neural Networks. *arXiv:1706.05098 [cs, stat]*
904 (2017).
- 905 58. T. L. Bank, E. A. Roth, P. Tinker, E. Granite, "Analysis of Rare Earth Elements in Geologic Samples
906 using Inductively Coupled Plasma Mass Spectrometry; US DOE Topical Report - DOE/NETL-
907 2016/1794" (NETL-PUB-20441, National Energy Technology Lab. (NETL), Pittsburgh, PA, (United
908 States), 2016); <https://doi.org/10.2172/1415779>.
- 909 59. P. T. Durdziński, M. Ben Haha, S. A. Bernal, N. De Belie, E. Gruyaert, B. Lothenbach, E. Menéndez
910 Méndez, J. L. Provis, A. Schöler, C. Stabler, Z. Tan, Y. Villagrán Zaccardi, A. Vollpracht, F.
911 Winnefeld, M. Zając, K. L. Scrivener, Outcomes of the RILEM round robin on degree of reaction of
912 slag and fly ash in blended cements. *Mater Struct* **50**, 135 (2017).

- 913 60. P. M. Granitto, C. Furlanello, F. Biasioli, F. Gasperi, Recursive feature elimination with random
914 forest for PTR-MS analysis of agroindustrial products. *Chemometrics and Intelligent Laboratory*
915 *Systems* **83**, 83–90 (2006).
- 916 61. M. Gevrey, I. Dimopoulos, S. Lek, Review and comparison of methods to study the contribution of
917 variables in artificial neural network models. *Ecological Modelling* **160**, 249–264 (2003).
- 918 62. Y. P. Liu, M. G. Wu, J. X. Qian, Predicting coal ash fusion temperature based on its chemical
919 composition using ACO-BP neural network. *Thermochimica Acta* **454**, 64–68 (2007).
- 920 63. W. Liang, G. Wang, X. Ning, J. Zhang, Y. Li, C. Jiang, N. Zhang, Application of BP neural network
921 to the prediction of coal ash melting characteristic temperature. *Fuel* **260**, 116324 (2020).
- 922 64. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimeshein,
923 L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B.
924 Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: An Imperative Style, High-Performance Deep Learning
925 Library. *Advances in Neural Information Processing Systems* **32** (2019).
- 926 65. T. Chen, C. Guestrin, “XGBoost: A Scalable Tree Boosting System” in *Proceedings of the 22nd ACM*
927 *SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016;
928 <http://arxiv.org/abs/1603.02754>), pp. 785–794.
- 929 66. S. Santurkar, D. Tsipras, A. Ilyas, A. Madry, How Does Batch Normalization Help Optimization?
930 *arXiv:1805.11604 [cs, stat]* (2019).
- 931 67. J. Howard, S. Gugger, *Deep Learning for Coders with Fastai and PyTorch* (O’Reilly Media, Inc., 2020).
- 932 68. J. Heaton, *Introduction to Neural Networks with Java* (Heaton Research, Inc., 2008).
- 933 69. T. Evgeniou, M. Pontil, “Regularized multi-task learning” in *Proceedings of the Tenth ACM SIGKDD*
934 *International Conference on Knowledge Discovery and Data Mining* (Association for Computing
935 Machinery, New York, NY, USA, 2004; <https://doi.org/10.1145/1014052.1014067>)*KDD ’04*, pp. 109–
936 117.
- 937 70. D. Xu, Y. Shi, I. W. Tsang, Y.-S. Ong, C. Gong, X. Shen, Survey on Multi-Output Learning. *IEEE*
938 *Transactions on Neural Networks and Learning Systems* **31**, 2409–2429 (2020).
- 939 71. T. Standley, A. Zamir, D. Chen, L. Guibas, J. Malik, S. Savarese, “Which Tasks Should Be Learned
940 Together in Multi-task Learning?” in *International Conference on Machine Learning* (PMLR, 2020;
941 <http://proceedings.mlr.press/v119/standley20a.html>), pp. 9120–9132.
- 942 72. C. Kuenneth, A. C. Rajan, H. Tran, L. Chen, C. Kim, R. Ramprasad, Polymer informatics with multi-
943 task learning. *Patterns* **2**, 100238 (2021).

- 944 73. M. Bettinelli, P. Taina, Rapid analysis of coal fly ash by x-ray fluorescence spectrometry. *X-Ray Spectrometry* **19**, 227–232 (1990).
945
- 946 74. R. Guo, L. Zhang, Y. Zheng, H. Yao, Accurate and stable measurement of ash in coal by X-ray
947 fluorescence spectrometry based on partial least squares. *Frontiers in Physics* **10** (2022).
- 948 75. S. M. Lundberg, S.-I. Lee, “A Unified Approach to Interpreting Model Predictions” in *Advances in*
949 *Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus,
950 S. Vishwanathan, R. Garnett, Eds. (Curran Associates, Inc., 2017; [http://papers.nips.cc/paper/7062-](http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf)
951 [a-unified-approach-to-interpreting-model-predictions.pdf](http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf)), pp. 4765–4774.
- 952 76. L. S. Shapley, A. E. Roth, *The Shapley Value: Essays in Honor of Lloyd S. Shapley* (Cambridge University
953 Press, 1988).
- 954 77. Y. Song, Y. Wang, K. Wang, G. Sant, M. Bauchy, Decoding the genome of cement by Gaussian
955 Process Regression. 7.
- 956 78. Materials informatics: From the atomic-level to the continuum - ScienceDirect.
957 [https://www.sciencedirect.com/science/article/pii/S1359645419300667?casa_token=XdT2H2s5rGk](https://www.sciencedirect.com/science/article/pii/S1359645419300667?casa_token=XdT2H2s5rGkAAAAA:gAKBFu6NFuVtrlHmqYQvtdKz-C27S-7MQjRdBuBDlelmM6ggms45WytJaNhlF3PHhXLtcmvD8tmT)
958 [AAAAA:gAKBFu6NFuVtrlHmqYQvtdKz-C27S-](https://www.sciencedirect.com/science/article/pii/S1359645419300667?casa_token=XdT2H2s5rGkAAAAA:gAKBFu6NFuVtrlHmqYQvtdKz-C27S-7MQjRdBuBDlelmM6ggms45WytJaNhlF3PHhXLtcmvD8tmT)
959 [7MQjRdBuBDlelmM6ggms45WytJaNhlF3PHhXLtcmvD8tmT](https://www.sciencedirect.com/science/article/pii/S1359645419300667?casa_token=XdT2H2s5rGkAAAAA:gAKBFu6NFuVtrlHmqYQvtdKz-C27S-7MQjRdBuBDlelmM6ggms45WytJaNhlF3PHhXLtcmvD8tmT).
- 960 79. B. Ouyang, Y. Song, Y. Li, F. Wu, H. Yu, Y. Wang, Z. Yin, X. Luo, G. Sant, M. Bauchy, Using machine
961 learning to predict concrete’s strength: learning from small datasets. *Eng. Res. Express* **3**, 015022
962 (2021).
- 963 80. Y. Zhou, F.-J. Chang, L.-C. Chang, I.-F. Kao, Y.-S. Wang, Explore a deep learning multi-output
964 neural network for regional multi-step-ahead air quality forecasts. *Journal of Cleaner Production* **209**,
965 134–145 (2019).
- 966 81. L. Melas-Kyriazi, “The Mathematical Foundations of Manifold Learning” (arXiv:2011.01307, arXiv,
967 2020); <https://doi.org/10.48550/arXiv.2011.01307>.
- 968 82. Z. Yang, Y. Li, Q. Lou, D. Liu, Y. Ning, S. Yang, Y. Tang, Y. Zhang, Z. Tang, X. Wang, Release of
969 uranium and other trace elements from coal ash by (NH₄)₂SO₄ activation of amorphous phase.
970 *Fuel* **239**, 774–785 (2019).
- 971 83. X. Querol, J. Fernández-Turiel, A. López-Soler, Trace elements in coal and their behaviour during
972 combustion in a large power station. *Fuel* **74**, 331–343 (1995).
- 973 84. Y. Sun, M. Wu, L. Zheng, B. Wang, Y. Wang, Uranium speciation in coal bottom ash investigated
974 via X-ray absorption fine structure and X-ray photoelectron spectra. *Journal of Environmental*
975 *Sciences* **74**, 88–94 (2018).

- 976 85. P. Catalfamo, S. Di Pasquale, F. Corigliano, L. Mavilia, Influence of the calcium content on the coal
977 fly ash features in some innovative applications. *Resources, Conservation and Recycling* **20**, 119–125
978 (1997).
- 979 86. B. Fu, J. C. Hower, W. Zhang, G. Luo, H. Hu, H. Yao, A review of rare earth elements and yttrium
980 in coal ash: Content, modes of occurrences, combustion behavior, and extraction methods. *Progress
981 in Energy and Combustion Science* **88**, 100954 (2022).
- 982 87. Z. Liu, C. T. Wu, M. Koishi, Transfer learning of deep material network for seamless structure–
983 property predictions. *Comput Mech* **64**, 451–465 (2019).
- 984 88. W. Franus, M. M. Wiatros-Motyka, M. Wdowin, Coal fly ash as a resource for rare earth elements.
985 *Environ Sci Pollut Res* **22**, 9464–9474 (2015).
- 986 89. Z. Wen, H. Liu, M. Zhou, C. Liu, C. Zhou, Explainable machine learning rapid approach to evaluate
987 coal ash content based on X-ray fluorescence. *Fuel* **332**, 125991 (2023).
- 988