ACS Sustainable Chemistry & Engineering

pubs.acs.org/journal/ascecg

Research Article

Machine Learning Enables Rapid Screening of Reactive Fly Ashes Based on Their Network Topology

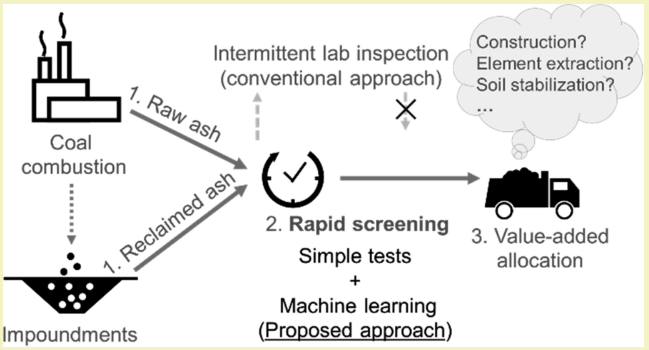
Yu Song, Kai Yang, Jingyi Chen, Kaixin Wang, Gaurav Sant,* and Mathieu Bauchy*

Cite This: *ACS Sustainable Chem. Eng.* 2021, 9, 2639–2650

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Fly ash, a byproduct of coal combustion, can be used as supplementary cementitious material (SCM) to replace ordinary portland cement (OPC) in concrete. This generates revenue for coal power plant operators and also reduces the CO_2 intensity of the binder fraction of a concrete (each ton of OPC replaced by fly ash results in 0.9 ton of avoided CO_2 emissions, if the fly ash is considered to have no carbon footprint). However, the use of fly ash in concrete has thus far been limited to replacement levels less than 20 mass % due to uncertainties in their performance as SCM. Although the ability of a fly ash to replace cement in concrete is largely determined by the reactivity of its amorphous phase, characterizing fly ashes' amorphous phase is complex and cost prohibitive, which has thus far prevented any high-throughput screening of fly ashes to assess their suitability as SCMs. Here, we introduce a machine-learning-based methodology that enables robust screening of reactive fly ashes based solely on fast, inexpensive bulk characterization (X-ray fluorescence: XRF), by using the network topology of fly ashes' amorphous phase as a structural proxy for their reactivity. On the basis of a data set of more than 100 fly ashes, we train an artificial neural network (ANN) model that offers accurate predictions of the mass fraction of fly ashes' amorphous phase and the network topology thereof. This new method seeks to maximize the beneficial use of fly ashes obtained from routine production, as well as to identify opportunities for the reclamation of ashes that are presently stored in impoundments.



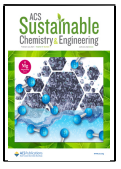
KEYWORDS: *Machine learning, Waste-to-resource, Fly ash recycling, Silicates reactivity, Topological constraint theory*

INTRODUCTION

Concrete, a mixture of ordinary portland cement (OPC), sand, stone, water, and chemical admixtures, is the most used material in the world after water.[1] Although concrete is a key enabler to modern society, the production of OPC (the binding phase of concrete) is responsible for nearly 10% of global CO_2 emissions.[2] The carbon footprint of concrete can be reduced by replacing OPC by supplementary cementitious materials (SCMs).[3] For reason of abundance, fly ash, a byproduct of coal combustion, is the most used SCM in concrete.[4] Fly ash takes the form of a heterogeneous, partially disordered aluminosilicate phase.[5] As a rough estimation, about 0.9 ton of CO_2 can be eliminated for each ton of cement being replaced by fly ash (if the fly ash is considered to have no carbon footprint).[6,7] In addition to reducing the carbon impact of concrete, the beneficial use of fly ash as SCM in concrete not only generates revenue for plant operators, but also reduces the fraction of fly ash that is disposed of in landfills and impoundments, which, in turn, reduces the risk of environmental damage due to leakage of contaminants into groundwater, release of contaminants into the air as dust, and failure of surface impoundments.[8] Furthermore, high-quality fly ashes can benefit concrete products, for example, by resulting in improved long-term strength, enhanced durability, and/or refined flowability.[4]

However, as the number of coal-based power plants decreases, especially in the U.S., so does the production of fly ashes.[9−11] As a result, there exists a growing interest in not only increasing the fraction of freshly produced fly ash that can be beneficially utilized in concrete, but also in isolating and reclaiming fly ashes that had previously been deposited in impoundments.[3] In both cases, fly ash's beneficial use is complicated by its substantial compositional variability, which results from variations in coal source, combustion protocols, pollution control operations, etc.[4] This is significant as the composition of fly ashes greatly affects their performance as SCM in concrete.[4,12,13] Hence, there is an urgent need for new

Received: September 21, 2020
Revised: January 29, 2021
Published: February 8, 2021



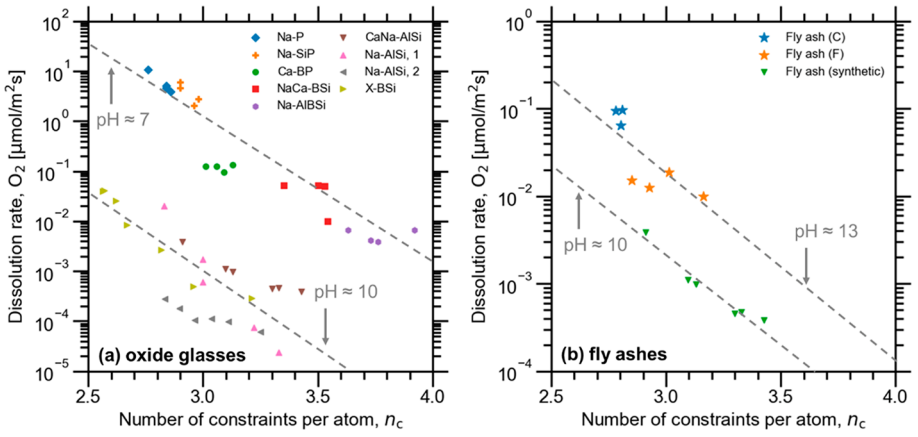


Figure 1. Correlation between the dissolution rate (D) and the number of constraints per atom (n_c) in (a) glass samples ranging over a wide variety of compositional families, that is, phosphate (Na−P), silicophosphate (Na−SiP), borophosphate (Ca−BP), borosilicate (NaCa−BSi), and aluminoborosilicate (Ba−AlBSi) glasses (solution pH at around 7),[38] and aluminosilicate (CaNa−AlSi, Na−AlSi) and borosilicate (X−BSi) glass (solution pH at around 10);[36,37,42−44] and (b) a selection of commercial (covering both Class C and Class F, pH 13) and synthetic (homogeneous glasses with compositions similar to those of fly ashes, pH 10) fly ash samples.[21,42] For a consistent comparison across the different glasses, the dissolution rate results are normalized in terms of dissolved O_2 (e.g., the dissolution of 1 mol of SiO_2 results in 1 mol of dissolved O_2). The dashed lines in both plots are exponential fits of the form $D = D_0 \cdot \exp(-E_0 \cdot n_c/RT)$, where D_0 is the dissolution attempt frequency in the absence of any constraint ($n_c = 0$), E_0 is the energy that is needed to break a unit constraint, R is the perfect gas constant, and T is the temperature.

rapid screening methodologies that can efficiently discriminate usable from nonusable fly ashes, both from fresh production and from impoundment reclamation. To this end, the most accepted classification method is specified in ASTM C618,[14] which is based on relatively simple measurements, that is, X-ray fluorescence (XRF), density measurement, and particle size analysis. From this standard, fly ashes can be roughly divided as "Class C" and "Class F" on the basis of their CaO content, wherein Class C ashes contain more than 18% CaO by mass and typically exhibit self-cementing properties. In contrast, Class F ashes usually consist of a higher sum of SiO_2, Al_2O_3, and Fe_2O_3 and tend to offer little self-cementing value, if any. However, many studies have emphasized that this specification is largely oversimplified and often too conservative to precisely screen fly ashes on the basis of their performance as SCM.[15−18] It is therefore expected that a vast, largely untapped amount of usable fly ashes is currently unnecessarily deposited in landfills on account of being non-C618-compliant.

Rather than being solely dictated by its CaO content as per C618, the ability of a fly ash to develop strength in concrete depends on its "reactivity". While the term "reactivity" is fairly vague and can be associated with various metrics and measurements,[19,20] the dissolution rate of SCMs in high-pH cementitious environments has been increasingly recognized as a reliable quantification of reactivity.[13,21−27] Specifically, because its crystalline compounds are usually fairly inert, the reactivity of a fly ash is usually determined by that of its amorphous phase, which accounts for 50−90% of fly ashes by mass.[28−30] The amorphous phases found in fly ash are compositionally and structurally similar to those of conventional alkali/alkaline earth aluminosilicate glasses.[13,21]

Importantly, recent studies have demonstrated that the dissolution rate of silicate glasses at high pH (as a proxy of the reactivity) is governed by the topology of their atomic network.[21,35,36] The network topology of glasses can elegantly be described by topological constraint theory (TCT), which simplifies their complex, disordered atomic structure into simpler mechanical trusses, where some nodes (the atoms) are connected to each other by some topological constraints (the chemical bonds).[31,32] Following Maxwell's stability criterion,[33] glasses can be classified as flexible, stressed-rigid, and isostatic when the number of topological constraints per atom constraint n_c is lower than, larger than, and equal to the number of degrees of freedom per atom (i.e., 3), respectively.[34]

As summarized in Figure 1a, several studies have demonstrated that the dissolution rate of a wide verity of oxide glasses (including aluminosilicate) is inversely correlated to n_c,[36−44] which establishes n_c as a simple, yet powerful structural metric that offers a convenient proxy for reactivity. In that regard, n_c has recently been shown to properly capture the reactivity of fly ashes at high-pH in a similar fashion. For example, Oey et al. demonstrated the existence of such a correlation using a group of commercial fly ashes (see Figure 1b).[21] It is generally observed that fly ashes featuring a flexible amorphous phase ($n_c < 3$, i.e., under-constrained) are notably more reactive than those presenting a stressed-rigid amorphous phase ($n_c > 3$, i.e., over-constrained).[21,42,43]

However, the enumeration of the topological constraints acting in the amorphous phase of fly ashes (i.e., to compute n_c) requires accurate knowledge of the composition and structure of the noncrystalline phase.[13,21] Although this information can be accessed via advanced characterization methods, such as quantitative X-ray diffraction (QXRD) and scanning electron microscopy with energy dispersive X-ray spectroscopy (SEM-EDS),[12,13,45−47] the complexity and cost associated with these techniques are not compatible with any systematic, rapid, high-throughput screening of fly ashes.

Nonetheless, previous studies have suggested that the composition of the amorphous phase in a fly ash is correlated (but not identical) to those of the bulk fly ash (i.e., as averaged over all of the crystalline and noncrystalline phases).[4,48] However, the physical and chemical nature of the mapping between the properties of the bulk fly ash and those of its amorphous phase remains unclear and is likely a complex manifestation of the features of the parent coal, the combustion thereof, as well as the thermal history of the fly ash.[4] Such linkages (if any) can be revealed by machine learning algorithms, which excel at uncovering hidden patterns

https://dx.doi.org/10.1021/acssuschemeng.0c06978
ACS Sustainable Chem. Eng. 2021, 9, 2639−2650

within data sets.[49] Over the past decade, advances in machine learning have unveiled new possibilities for predicting material properties as a function of structural or compositional fingerprints,[50] including for silicate glasses.[51] In that regard, some recent studies have applied machine learning to predict the influence of fly ash on the compressive strength and durability of concrete,[52−55] but little attention has been paid to explore the possibility of using machine learning to characterize the composition and structure of fly ashes.

In this work, we propose a topology-informed machine learning approach to screen reactive fly ashes solely on the basis of their bulk chemical composition, that is, which can quickly and easily be measured by XRF. The XRF analysis that serves as input to the methodology developed herein can be conducted on-site, in real-time, and in a cost-efficient manner. A machine learning model then maps the bulk XRF composition of a fly ash to (i) the mass fraction of its amorphous phase and (ii) the network topology thereof (as captured by n_c, i.e., a proxy for fly ashes' reactivity). This machine learning model is trained and tested on the basis of a curated data set comprising more than 100 fly ashes with varying compositions and reactivities. As illustrated in Figure 2,

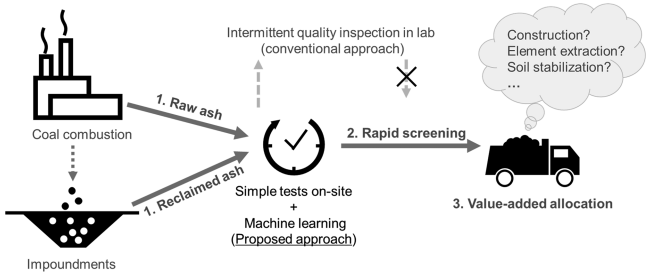


Figure 2. A schematic of the proposed machine-learning-based screening approach. Easily measurable fly ash properties are mapped to key performance metrics to inform decision-making regarding which end usage should be preferred.

this approach enables the rapid screening of fly ashes, because the most optimal usage of fly ashes can be meaningfully determined on the basis of their predicted reactivity. Our approach combines recent advances in cutting-edge artificial intelligence (AI) methods and glass science knowledge to support the sustainable management of industrial waste, which showcases a novel pathway to develop green engineering.

METHODOLOGY

The Fly Ash Data Set. Data are vital because they are the "textbook" from which a machine learning model learns. The fly ash data set used herein (detailed in section S9) is collected from 15 publications, and it is comprised of 107 fly ash samples covering a large variety of fly ash compositions from diversified sources and regions. We solely extract from the sourced literature the fly ash samples for which both the mass fraction of the amorphous phase and the composition thereof (in terms of the fraction of the major oxides) are reported. Although the fraction of the amorphous phase can be measured via different methods (e.g., with QXRD and SEM-EDX), we do not observe any systematic deviations across the literature, so that we expect that this inconsistency does not notably affect the training of the model. A brief summary of the characteristics of the fly ash samples used in this study, together with all relevant references, can be found in section S1 of the Supporting Information.

Because we aim to develop a fly ash screening platform that solely relies on easily measurable features, the model inputs are selected from the parameters specified by ASTM C618.[14] Inputs are comprised of the bulk mass fraction (obtained by XRF) of the six major oxides that are systematically measured and reported in all of the sourced publications: (1) CaO, (2) Al_2O_3, (3) SiO_2, (4) Fe_2O_3, (5) MgO, and (6) $Na_2O + 0.658K_2O$ (i.e., total alkali content). As a reference, the statistical distributions of the six adopted features are summarized in Table 1. On average, the input oxides considered herein sum to 94%. All of the 107 samples are compliant with the ASTM C618 requirement; that is, the sum of the SiO_2, Al_2O_3, and Fe_2O_3 fractions exceeds 50%.[14] Importantly, it can be seen from Table 1 that the input features cover a wide compositional envelope that ranges over both Class C and Class F fly ashes.[5] In contrast, other minor components such as loss on ignition (LOI), SO_3, and fineness are omitted because (i) they are not systematically measured and reported throughout all of the publications considered herein and/or (ii) they do not notably affect the rigidity of the atomic network of the amorphous phase of fly ashes. To quantify the influence of excluding minor oxides on the accuracy of our model, we retrained our model while accounting for these minor components (i.e., SO_3 and LOI), where the missing values were imputed on the basis of the average of their five nearest neighbors (see section S6 of Supporting Information). We observe a consistent decrease in the model accuracy when accounting for these minor components. This is possibly caused by the increased dimensionality of the input features (which requires a large volume of samples to train the model) and/or the fact that some of the data points presenting missing values do not have close neighbors, which makes it difficult to meaningfully use data imputation techniques.

Regarding the model outputs (i.e., the target properties that the models are trained to predict), we consider three major parameters that influence fly ash reactivity: (i) the mass fraction of the amorphous phase in the fly ash, (ii) the chemical composition thereof, and (iii) the number of topological constraints per atom (n_c, hereafter, number of constraints) within the amorphous phase. The first two outputs (fraction of the amorphous phase and composition

Table 1. Variations of the Six Selected Input Features for the 107 Fly Ash Samples Sourced from the Literature

distribution statistics		chemical composition from XRF (mass %)					
		Al_2O_3	CaO	Fe_2O_3	SiO_2	MgO	$Na_2O + 0.658K_2O$
percentile	0th	0.1	0.1	0.1	27.1	0.2	0.0
	20th	18.2	1.8	4.5	41.5	0.9	1.1
	40th	20.7	4.7	5.7	48.5	1.5	1.5
	60th	25.0	9.9	6.4	53.9	2.3	1.9
	80th	28.3	20.2	7.9	56.6	4.1	2.5
	100th	35.6	38.9	16.0	97.0	10.8	8.7
median		23.0	8.3	6.0	51.3	2.0	1.7
mean		22.6	10.2	6.2	50.4	2.5	2.0
standard deviation		6.5	9.3	2.7	12.2	1.9	1.3

https://dx.doi.org/10.1021/acssuschemeng.0c06978
ACS Sustainable Chem. Eng. 2021, 9, 2639−2650

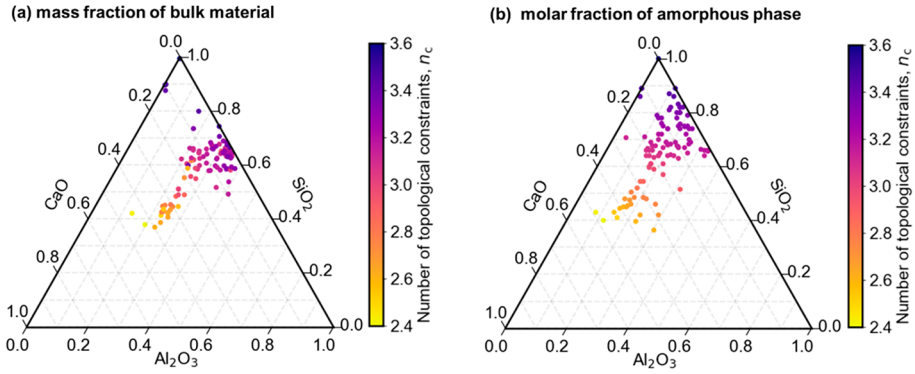


Figure 3. Ternary plot illustrating (a) the bulk XRF composition (in terms of mass fractions) and (b) the composition of the amorphous phase (in terms of molar fraction) of the fly ashes considered in this study. All compositions are expressed and normalized within the $CaO-Al_2O_3-SiO_2$ ternary system. Each data point is colored on the basis of the associated number of topological constraints per atom.

thereof) are directly reported in the literature. In contrast, the third output (number of constraints) is derived via a common constraint enumeration on the basis of the knowledge of the composition of the amorphous phase.[34] To estimate n_c for each fly ash sample, we simplify the amorphous phases by solely focusing on the most common and dominant oxides (i.e., CaO, SiO_2, and Al_2O_3), while filtering out less common and minor oxides that have a smaller influence on n_c.[21] This simplification is necessary to ensure consistency between all of the fly ash samples in the data set. Herein, the number of topological constraints per atom is calculated on the basis of the following formula:

$$n_c = \frac{11 - 12x + y}{3 - 2x + 2y} \quad (1)$$

where x and y are the CaO and Al_2O_3 molar fractions in the amorphous phase, respectively. More details about the constraints enumeration are available in section S2 of the Supporting Information. The full data set is also provided in section S9.

To further illustrate the compositional variety of the 107 fly ash samples present in the data set, each of them is mapped onto the $CaO-Al_2O_3-SiO_2$ (CAS) ternary diagrams in Figure 3. For illustration purposes, all compositions are here normalized such that the sum of the CaO, Al_2O_3, and SiO_2 oxides fractions equals 100%. Figure 3a shows the bulk XRF composition of the fly ashes in the data (i.e., the inputs of the present analysis). Note that these compositions are here expressed in mass fractions, because such metrics are commonly used in ASTM C618. We note that the fly ashes considered herein range over both the Class F and the Class C families, albeit with excess in Class F representation. For comparison, Figure 3b shows the composition of the amorphous phase within the same fly ashes (i.e., one of the outputs of the present analysis). Note that these compositions are here expressed in molar fractions, because the knowledge of the molar fractions of each oxide is a prerequisite to calculating the number of topological constraints per atom (see eq 1). Overall, although we observe some level of correlation between bulk and amorphous phase compositions, no direct, obvious transformation/correlation rule exists.

Neural Network Model Developed in This Study. Herein, we adopt the neural network approach to predict fly ash reactivity. Indeed, on account of its flexibility, this machine learning method is expected to offer an ideal path to map the XRF composition of fly ashes to properties of their amorphous phase, because this relationship is ill-defined and its nature (and analytical formulation, if any) is currently unknown. For the reader's reference, we detail the key concepts of neural networks (and of the training and evaluation thereof) in section S3 of the Supporting Information.

The accuracy of a neural network depends on various factors, such as the volume of the data set, number of the input variables, complexity of the learning task, etc. In general, a neural network that is not complex enough cannot properly interpolate the true correlation between inputs and output. In turn, complex neural networks (e.g., comprised of many neurons and/or layers) tend to require a high volume of data to be properly trained (higher than the amount of data needed to train a simpler model). Here, on the one hand, a single layer of hidden neurons is unlikely to accurately capture the complex relationship between the bulk XRF composition of the fly ash and the properties of the amorphous phase. On the other hand, the data set used herein is relatively small, but only presents a few input features. As a tentative rule of thumb, a neural network model should be roughly comprised of (i) a first hidden layer with a number of neurons that is 1−2 times larger than the number of inputs and (ii) subsequent hidden layer(s) with progressively decreasing numbers of neurons.[56] On the basis of this rationale, we adopt a two-layer neural network using the multilayer perceptron regressor in scikit-learn.[49] Furthermore, the number of neurons in the two layers is adjusted to maximize the model accuracy. Eventually, we adopt 10 and 3 neurons in the first and second hidden layers, respectively. We use the Rectified Linear Unit (ReLU) as activation function. We adopt the Limited-memory Broyden−Fletcher−Goldfarb−Shanno (L-BFGS) algorithm as optimizer, because it has been shown to efficiently handle small data sets.[57] The model accuracy is evaluated on the basis of the coefficient of determination (R^2) and the mean absolute percentage error (MAPE). The remaining model settings are fixed on the basis of default values in scikit-learn.

Here, two hyperparameters are tuned for model optimization. The first one is the number of training epochs (i.e., the total number of rounds of training, that is, the number of times the model is exposed to the entire training set to update the parameters within each neuron). The second is an L_2 regularization term, alpha,[58] which needs to be adjusted to prevent the model from under- or over-fitting the data set. The hyperparameter tuning process is carried out in a two-step fashion. First, the effect of various numbers of epochs (up to 2000) is investigated while fixing the regularization parameter alpha as its default value (i.e., 0.0001). After the optimal number of epochs is determined, different alpha values ranging between 0.00001 and 10 are compared. During this step, the model is expected to transit from the under-fitting (at high alpha) to over-fitting (at low alpha) regimes, in between which an optimal point can be found, that is, wherein the accuracy of the validation set is maximized.

Here, 85% of the fly ash data set (i.e., 91 out of the 107 samples) is selected for the training phase. These 91 samples are further divided into 85% (77 samples) for training the model under different hyperparameters (i.e., training set) and 15% (16 samples) for evaluating the corresponding model accuracy (i.e., validation set). To obtain a more statistically meaningful assessment of the model performance, each model (with given hyperparameter settings) is evaluated on the basis of 10 repetition trials, wherein the training and validation split is updated in each trial.

Because of the limited data set size, we find that, when the validation/test set is sampled randomly, there is a chance that its data distribution does not echo that of the training set (e.g., the extraction

https://dx.doi.org/10.1021/acssuschemeng.0c06978
ACS Sustainable Chem. Eng. 2021, 9, 2639−2650

can be concentrated at one end of the entire data distribution). As a result, the model training may be statistically biased, so that the model optimization would be inefficient. To address this problem, we further implement stratified sampling to boost the model training.[50] To ensure more statistically representative sampling, this approach first divides the entire data set into several subsets and then extracts samples randomly from each subset by proportion. In addition, we conduct a SHapley Additive exPlanations (SHAP) analysis to evaluate the influence of each input feature to the model prediction.[59] More details regarding stratified sampling and the SHAP analysis are provided in sections S4 and S7 of the Supporting Information, respectively.

RESULTS

Prediction of the Mass Fraction of the Amorphous Phase in Fly Ashes. The optimization of the model hyperparameters/architecture is presented in section S5 of the Supporting Information. Once optimized, the developed models are then trained with stratification (see section S4 of the Supporting Information) and subsequently used to, first, predict the mass fraction of the amorphous phase within the fly ashes (see Figure 4). As suggested by the even distribution of

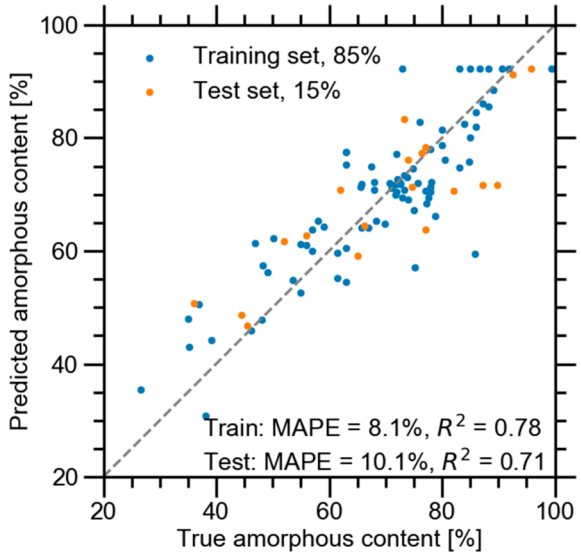


Figure 4. Predicted versus measured amorphous content (mass %) of the fly ashes considered herein. The training set used herein comprises 85% of the whole fly ash data set (as merged from the training and validation sets previously used in the training/validation phase), while the remaining 15% of the data set is used as the test set (which is kept fully hidden from the model during the training phase). The $y = x$ dashed line indicates perfect agreement.

the data points along the line of equality $y = x$, we find that the neural network exhibits a fairly good accuracy in predicting the amorphous content of unknown fly ashes. On the basis of the R^2 and MAPE metrics, the model is found to achieve a slightly higher accuracy on the training set than on the test set. However, no distinct systematic bias can be observed on either set. To our knowledge, this is the first time that a model accurately predicts the mass fraction of the amorphous phase of fly ashes while solely relying on easily measurable features (i.e., bulk XRF composition). Such accuracy (i.e., MAPE = 10%) is considered to be satisfactory considering that our model does not explicitly consider details of the fly ash quenching process (e.g., processing temperature, cooling rate, etc.), which may affect the degree of crystallinity of the ashes.[4]

Prediction of the Reactivity of the Fly Ashes. Next, we focus on predicting the number of constraints per atom in the amorphous phase of the fly ashes. This reduced-dimensionality metric captures the rigidity of the atomic network[31,35] and, thereby, can be used to predict the reactivity of fly ashes; flexible ($n_c < 3$) and stressed-rigid ($n_c > 3$) phases tend to exhibit high and low reactivity, respectively.[21,42,60] To this end, we use the same six input features as the model presented in the previous subsection. During our initial exploration, we attempted to use the amorphous content predicted by the model presented in the previous subsection as an additional input for the model predicting n_c. However, that approach did not yield any notable improvement in accuracy and, therefore, is not considered in the following.

In the following, we conduct and compare two distinct approaches (indirect and direct prediction) that both ultimately aim to predict n_c. On the one hand, in the case of the indirect prediction approach, we train three distinct neural networks that are specifically used to predict the CaO, Al_2O_3, and SiO_2 molar fractions in the fly ash amorphous phase. These predicted molar fractions are then subsequently used as inputs to analytically calculate the number of constraints per atom following eq 1. On the other hand, the second approach simply considers n_c as the direct output of the neural network. This comparison is intended to assess whether the neural network can autonomously learn the relationship between composition and network rigidity captured by eq 1.

We first consider the indirect method. Figure 5 illustrates the accuracy of the three distinct models that aim to independently

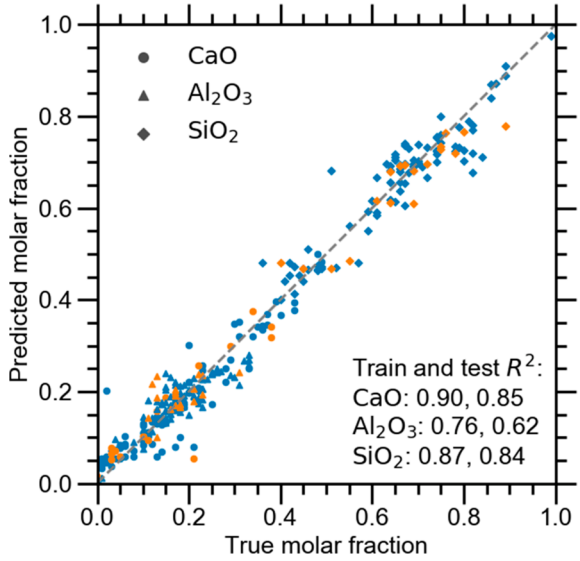


Figure 5. Predicted versus measured CaO, Al_2O_3, and SiO_2 molar fractions within the fly ashes' amorphous phase. The training set comprises 85% of the whole fly ash data set (blue symbols), while the remaining 15% of the data is used as the test set (orange symbols). The $y = x$ dashed line indicates perfect agreement.

predict the CaO, Al_2O_3, and SiO_2 molar fractions in the fly ash amorphous phase. Overall, we find that these models exhibit a satisfactory accuracy, for both the training and the test sets. Among the three oxides, the fact that Al_2O_3 is associated with a lower accuracy might be on account of the smaller variation of this feature in the fly ash data set (see Table 1). Importantly, the model predicting the CaO molar fraction exhibits the highest accuracy, which is notable considering the key

https://dx.doi.org/10.1021/acssuschemeng.0c06978
ACS Sustainable Chem. Eng. 2021, 9, 2639−2650

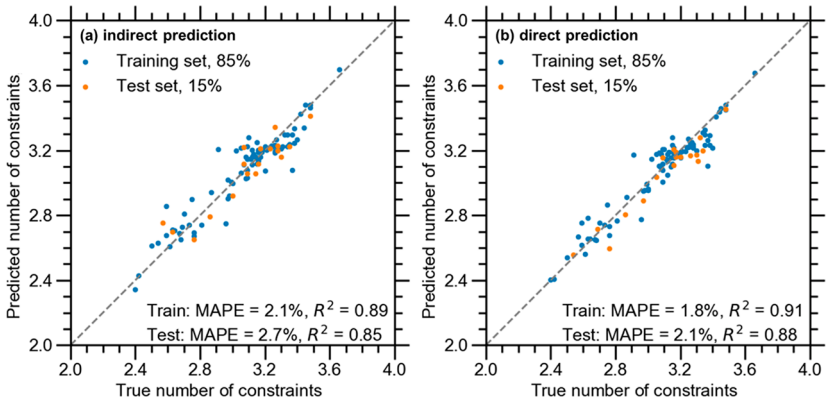


Figure 6. Predicted versus measured number of topological constraints per atom (n_c) in the amorphous phase of the fly ashes considered herein, wherein n_c is obtained by (a) indirect prediction, that is, wherein n_c is calculated on the basis of the composition of the amorphous phase predicted by the artificial neural network model presented in Figure 5, and (b) direct prediction of a single artificial neural network model (wherein n_c is the output neuron). In both approaches, the models are trained on the basis of 85% of the whole data set and subsequently tested with the remaining 15%. The $y = x$ dashed line indicates perfect agreement.

Table 2. Confusion Matrix and Intersection over Union (IoU) Classification Accuracy Offered by Our Neural-Network-Based Classification Model Aiming to Screen Reactive from Nonreactive Fly Ashes[a]

		ASTM C618 estimation[b]			neural network prediction[c]		
confusion matrix		highly reactive (Class C)	poorly reactive (Class F)	IoU accuracy	highly reactive ($n_c < 3$)	poorly reactive ($n_c > 3$)	IoU accuracy
true result	highly reactive	22	8	71%	train: 24.6 test: 5.6	train: 1.6 test: 0.4	train: 92% test: 90%
	poorly reactive	1	76	89%	train: 0.4 test: 0.2	train: 61.4 test: 13.0	train: 97% test: 96%

[a]The accuracy of our model is compared to that offered by the conventional ASTM C618 standard.[14] [b]In ASTM C618, reactive fly ashes (Class C) have >18% CaO by mass in the bulk material. [c]The results are averaged from five independent trials.

importance of this oxide in discriminating reactive from nonreactive fly ashes.[30] Figure 6a then illustrates the ability of this indirect approach to offer a realistic prediction of the number of constraints per atom n_c. Overall, we find that this indirect method exhibits a robust prediction across the whole range of n_c values between 2.4 (flexible) and 3.6 (stressed-rigid). This approach achieves a MAPE of 2.7% and an R^2 of 0.85 on the test set. As such, we find that this model is more accurate than that predicting the amorphous content (see Figure 4). This is not surprising because the number of constraints per atom is a quantity that is averaged over the entire amorphous phase, while, experimentally, the mass fraction of the amorphous phase is indirectly obtained by subtracting the total mass by the individual masses of each crystalline phase,[12] thereby resulting in larger measurement uncertainties.

Next, we focus on the direct prediction method. Figure 6b illustrates the ability of the direct approach to predict the number of constraints per atom n_c. Interestingly, we observe that the direct model exhibits an accuracy that is higher than that of the indirect approach. This is somewhat surprising because the models trained within the indirect approach are exposed to more information (i.e., the molar composition of the amorphous phase) than the model used in the direct approach, because each fly ash sample is associated with three distinct molar concentrations (CaO, Al_2O_3, and SiO_2), but only one number of constraints per atom n_c. As such, it appears that the indirect approach does not positively leverage this increased amount of information. The fact that the direct method offers an improved accuracy might be due to the fact that the indirect approach is negatively affected by the accumulation of prediction errors associated with each of the individual molar fractions. Overall, by reducing the dimensionality of the output, the direct model offers an improved accuracy and manages to successfully (i) map bulk XRF composition to amorphous phase composition and (ii) further convert the amorphous phase composition to its number of constraints per atom within a single model.

Screening of Reactive versus Nonreactive Fly Ashes. On the basis of the direct model presented in the previous subsection, we now assess the ability of our approach to robustly screen reactive from nonreactive fly ashes. To this end, to quantify the performance of our model, we categorize the fly ash samples contained in the present data set into two classes on the basis of their ground-truth state of atomic rigidity, which captures the propensity to be (i) highly reactive ($n_c < 3$) and (ii) poorly reactive ($n_c > 3$).

To establish a baseline, we first assess the ability of the ASTM C618 approach to properly classify the fly ashes present in the data set on the basis of whether they are classified as Class C (CaO mass fraction larger than 18%) or as Class F (CaO mass fraction lower than 18%), wherein Class C fly ashes are predicted to be reactive on the basis of ASTM C618.[14] We consider the intersection over union (IoU) metric to assess the classification accuracy (see Table 2). Notably, we find that the empirical estimation performed by ASTM C618 classifies reactive fly ashes with only a 71% accuracy. The error is associated with the fact that eight reactive fly ash samples (8 out of 30) are predicted to be poorly reactive on the basis of their fairly low CaO mass fraction, but are actually reactive on

https://dx.doi.org/10.1021/acssuschemeng.0c06978
ACS Sustainable Chem. Eng. 2021, 9, 2639−2650

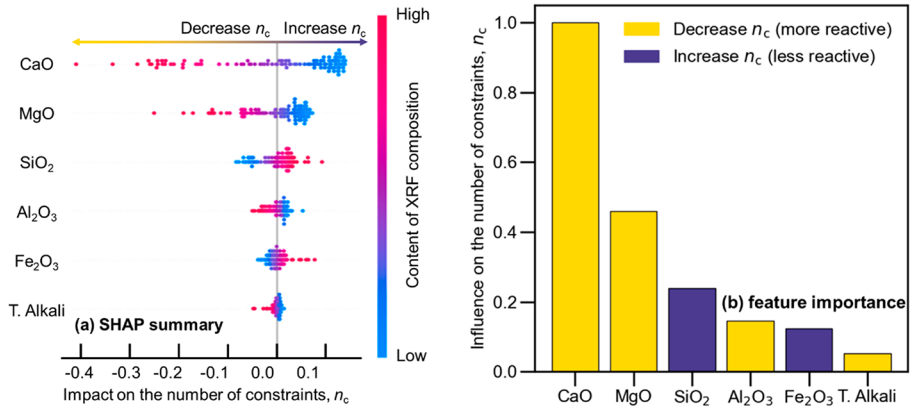


Figure 7. (a) Summary of the SHAP analysis showing the impact of the six input features (XRF bulk compositions) on the number of constraints per atom n_c predicted by the direct neural network model presented in Figure 6b and (b) normalized influence of the different input features.

account of the relative contents of higher Al_2O_3 and lower SiO_2 (more results regarding this are presented in the next subsection). These results highlight some potential limitations of ASTM C618, which are associated with its incomplete and largely empirical description of fly ash reactivity. More specifically, these results suggest a tendency for ASTM C618 to underestimate the reactivity of fly ashes because it predicts a non-negligible portion of reactive fly ashes to be nonreactive, while, in turn, it exhibits a very low propensity to misclassify nonreactive fly ashes as reactive.

We then assess the classification accuracy offered by the direct model presented in Figure 6b, wherein fly ash samples are directly classified as reactive or nonreactive on the basis of the n_c value predicted by the neural network model. Here, for statistical purposes, the accuracy of our model is determined on the basis of the average of five different models trained based on distinct training-test splits. Remarkably, we find that our direct neural network-based classification approach offers a notably improved classification accuracy as compared to ASTM C618. In detail, our model successfully classifies reactive fly ashes that exhibit low atomic rigidity with an accuracy of 92% and 90% on the training and test sets, respectively, and, in turn, classifies nonreactive fly ashes that exhibit high atomic rigidity with even higher accuracy of 97% and 96% on the training and test sets, respectively. These results indicate that, despite the limited size of the present data set, our machine learning model can discriminate reactive versus nonreactive fly ashes. Indeed, our approach exhibits higher precision than ASTM C618 as our approach exhibits a lower chance of misclassifying nonreactive fly ashes as reactive. Such precision is important to ensure that nonreactive fly ashes are properly utilized, because replacing cement by nonreactive fly ashes would otherwise hinder the performance of the concrete that is produced. More importantly, our approach features a higher recall when identifying reactive fly ashes, where the IoU accuracy is increased by about one-third (from 71% to 90%) as compared to ASTM C618. This highlights the fact that, on account of its increased accuracy, our approach is less conservative than ASTM C618 in classifying reactive fly ashes. This is important as it implies that a notable amount of reactive fly ashes that are presently estimated as being noncompliant based on ASTM C618 is actually reactive and could in fact be beneficially used to replace cement in concrete. As such, on account of its increased precision and recall, our machine learning approach establishes a promising option to screen as-produced fresh fly ashes in the plant, but also reclaim fly ashes that are presently stored in impoundments and that have previously been judged as noncompliant by ASTM C618, thereby creating value for these fly ashes that would otherwise be associated with an impoundment management cost.

Interpretation of the Artificial Neural Network Model. As an attempt to interpret the trained neural network model (which is known to be a "black box" in mapping inputs to output), we first conduct a SHAP analysis on the direct model (see Figure 6b) that predicts the number of constraints per atom n_c in the amorphous phase of the fly ashes. The SHAP analysis is a technique that is used to interpret the trends that have been learned by a trained machine learning model, based on quantifying the marginal contribution of the model inputs to the model prediction.[61] Here, it allows us to infer the importance of the individual input features (i.e., the XRF bulk compositions) in influencing the output number of constraints per atom. Importantly, unlike simpler correlation analysis, SHAP takes into account the interaction between variables (wherein a given variable may be weakly influential on its own, but highly influential when combined with another feature). Figure 7a provides a summary of the feature-specific distribution of the Shapley values (i.e., the marginal contribution to n_c, which is either positive or negative). In this plot, the y-axis refers to each input feature (as ranked in order of decreasing importance from top to bottom). The x-axis then refers to the Shapley value, which indicates how much is the model output (n_c) influenced by the feature; wherein the horizontal point dispersion indicates the range of feature impact, the vertical width represents the distribution of the data points in the original data set, and the color refers to the value of the input feature. As such, this analysis offers a direct access to the impact of each feature and to how this impact depends on the actual value of the feature. For example, Figure 7a indicates that, while, overall, the CaO XRF content negatively impacts the number of constraints, this effect is more pronounced at high CaO content. In contrast, high SiO_2 XRF content positively impacts the number of constraints.

On the basis of the summary plot of the SHAP analysis (Figure 7a), we further compute the overall feature importance (see Figure 7b), wherein the feature importance is obtained by first calculating the sum of the absolute marginal contribution of the individual features in all of the samples (so as to quantify the total impact), and then normalizing the total impact of each feature with that of the most influential one (in this study,

https://dx.doi.org/10.1021/acssuschemeng.0c06978
ACS Sustainable Chem. Eng. 2021, 9, 2639−2650

the CaO content). Generally speaking, these results match with expected trends: (i) network-forming elements (e.g., Si) tend to enhance the connectivity of the atomic network, thereby increasing the number of constraints per atom, while, in contrast, (ii) network-modifying elements (e.g., Ca, Mg, alkali) tend to depolymerize the silicate network, which, in turn, tend to decrease the number of constraints per atom.[28] In detail, the fact that CaO is by far the most influential feature concurs with the widely accepted fly ash Class C/F classification defined in ASTM C618.[14] Despite having a lower average concentration, MgO is found to be the second most influential feature. This echoes the fact that Ca and Mg alkaline earth create fairly similar low-energy ionic bonds with oxygen atoms and have a similar depolymerizing effect on the atomic network of silicate glasses. Thus, the presence of CaO and MgO in the amorphous phase of fly ashes tends to enhance their reactivity by lowering n_c. We find that SiO_2 ranks as the third most influential feature and, in contrast to CaO and MgO, tends to increase the number of constraints per atom (and, thereby, decrease fly ash reactivity). This is a consequence of the fact that Si is the base element forming the atomic backbone of the amorphous phase of fly ashes. As a network former, Si atoms create strong, directional ionocovalent bonds with oxygen neighbors, which tends to increase the number of both bond-stretching and bond-bending constraints in the network. Similarly, the fact that Fe_2O_3 tends to increase the number of constraints is in agreement with the fact that Fe is also usually considered as a network-forming species.[28] The role of Al_2O_3 is generally less clear, because, when combined with Si atoms, Al atoms are considered to be network formers in Al-poor peralkaline silicate glasses, but eventually exhibit a network-modifying role in Al-rich peraluminous silicate glasses.[28,48,60,62] The ambivalent role of Al_2O_3 likely explains its relatively low ranking in Figure 7b. However, the results in Figure 7 indeed suggest that the presence of Al_2O_3 in the fly ashes considered herein tends to mildly increase their reactivity. Finally, we find that the total alkali content presents the lowest influence on n_c. Note that, however, this observation does not imply that the total alkali content does not influence the reactivity of fly ash. Rather, it only implies that this feature is less influential than the other features investigated herein.

As a final attempt to interpret the predictions offered by the direct model aiming to predict the number of constraints per atom (presented in Figure 6b), we explore the prediction offered by this model over the entire calcium aluminosilicate (CAS) ternary system. To this end, we first define a typical reference fly ash XRF bulk composition, which is given by the XRF composition of the fly ash sample that is associated with the median n_c value over the whole fly ash data set. This yields the following reference composition: CaO (10.2%), Al_2O_3 (22.6%), SiO_2 (50.4%), MgO (2.5%), Fe_2O_3 (6.2%), and total alkali (2.0%). Starting from this reference composition, we then systematically vary the relative CaO, Al_2O_3, and SiO_2 XRF mass fractions by increments of 0.1%. This approach allows us to ensure that the targeted input XRF compositions are representative of real fly ashes comprised in our data set.

Figure 8 shows a ternary map illustrating the number of constraints per atom that is predicted over the entire CAS ternary compositional domain, wherein fly ashes are described in terms of their normalized relative mass fractions of CaO, Al_2O_3, and SiO_2. Figure 8 also shows the actual fly ash samples as superimposed points (i.e., the same ones as presented in Figure 3a). It is worth noting that the actual and virtual fly

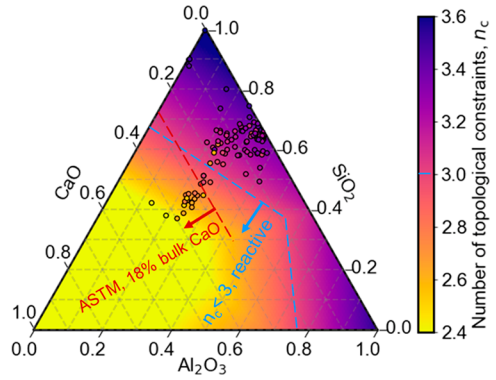


Figure 8. Ternary plot illustrating the number of topological constraints per atom (n_c) predicted by the artificial neural network model presented in Figure 6b as a function of the fly ash bulk XRF composition (mass fractions). The compositions of the real fly ashes present in the data set (and associated n_c values) are added for comparison. All compositions are expressed and normalized within the $CaO-Al_2O_3-SiO_2$ (CAS) ternary system. The dashed blue line indicates the classification boundary between highly reactive ($n_c < 3$) and poorly reactive ($n_c > 3$) fly ashes. The dashed red line indicates the equivalent ASTM classification boundary for reactive fly ashes (mass fraction of CaO exceeds 18% in the bulk material; here, the boundary is converted so as reflect the renormalization of the fly ash composition within the CAS ternary by fitting the present fly ash data set; see section S8 of the Supporting Information).

ashes compositions indicated on this map may not have exactly the same XRF compositions (due to differences in MgO, Fe_2O_3, and total alkali content between the actual samples and the assumed inputs). We first observe that the model offers a smooth evolution of n_c over the entire compositional domain (see Figure 8), which a posteriori confirms that the model is not over-fitted. Further, on the basis of this ternary map, we observe that, as expected, the fly ash reactivity is strongly affected by the CaO mass fraction, which manifests itself by the fact that the gradient of n_c is largely orthogonal to the direction of the CaO axis.

This analysis is then used to determine the boundary that separates reactive from nonreactive fly ashes within the $CaO-Al_2O_3-SiO_2$ (CAS) system (i.e., the hyperplane associated with $n_c = 3$, see Figure 8), which is then compared to the threshold offered by ASTM C618. After renormalizing the fly ash compositions within the CAS system, we find that the 18% of XRF content of CaO (i.e., the threshold designated by ASTM C618) is associated with a higher CaO content of about 27% within the CAS ternary system, on average. Details on this conversion are provided in section S8 of the Supporting Information. As this normalized threshold broadly holds for Figure 8, we observe that several fly ash samples are classified as reactive by the neural network model, but nevertheless predicted as nonreactive by ASTM C618. Note that the n_c boundary shown in Figure 8 represents only one projection of the $n_c = 3$ hyperplane (which varies as a function of the fractions of MgO, Fe_2O_3, and total alkali content). The contrast between the two approaches observed herein does not conflict with Table 2. Indeed, a large compositional domain of fly ashes can be reactive ($n_c < 3$) while presenting a CaO mass fraction that is lower than the 18% (as prescribed from ASTM C618), especially upon increasing Al_2O_3 mass fraction. This illustrates once again that ASTM C618 may present a propensity to underestimate fly ashes' reactivity. Note that one should be careful when interpreting the predictions of the

https://dx.doi.org/10.1021/acssuschemeng.0c06978
ACS Sustainable Chem. Eng. 2021, 9, 2639−2650

neural network model when it is extrapolated (i.e., when the model is interrogated for XRF compositions that are far from the compositions of the actual fly ash samples on which the model was trained). Nevertheless, this ternary map generally echoes with the expected compositional dependence of n_c in CAS glasses.[37] In addition, this map offers useful insights on the best possible route to remediate fly ashes, that is, to adjust the composition of initially nonreactive fly ashes to make them reactive.

DISCUSSION

The results herein demonstrate that, while using the knowledge of the bulk XRF composition of the fly ashes as the sole inputs, our neural network model is able to successfully predict the mass fraction of the amorphous phase, the composition thereof, and the number of constraints per atom therein (i.e., a proxy for reactivity). Despite the limited number of data points for training, our models are able to achieve a good accuracy on both the training and the test sets, which suggests that the size of the present data set is large enough to support the establishment of our screening models. Indeed, the performance of these models is fairly surprising as these relationships are somewhat mathematically ill-defined in the first place. This is because, although one can analytically calculate the average bulk composition of a fly ash on the basis of the knowledge of the composition and mass fraction of each individual phase (amorphous or crystalline), the inverse problem (i.e., predicting the fraction and composition of the individual phases on the basis of the knowledge of the average) does not have a unique solution. In other words, mathematically speaking, two fly ashes with similar bulk compositions can nevertheless exhibit differing compositions of their amorphous phase, because variations in the composition of the amorphous phase can be compensated by variations in the fractions of the other crystalline phases.

Nevertheless, the fact that the present model can successfully map XRF bulk composition to the fraction, composition, and rigidity of the amorphous phase suggests that some implicit correlations do indeed exist between these features within fly ashes, although the origin and nature of such correlations might not be readily capturable by intuition. In that regard, previous studies have offered some possible explanations for this kind of correlation.[4,48,63] Most likely, fly ashes with fairly similar XRF bulk compositions tend to exhibit comparable degrees of crystallinity (i.e., similar amorphous phase fraction), irrespective of other second-order conditions that are not explicitly captured as features within our data set. This is a consequence of the fact that, at a constant cooling rate, the glass-forming ability of a melt (i.e., its relative propensity to form a glass or a crystal upon cooling) strongly depends on the composition of the melt.[64,65] Further, although there might not exist a direct causal relationship between the average bulk XRF composition and that of the amorphous phase, these two features may be two distinct consequences of the same causes, that is, the origin of the coal, combustion quality, cooling protocol, etc. More generally, the mapping accuracy that is achieved by the model presented herein is a manifestation of the fact that when describing a phenomenon based on multiple features (i.e., in multiple dimensions), groups of observations that are close to each other along some selected dimensions also tend to be at the vicinity of each other along other dimensions.[66,67] This is a clear advantage of machine learning approaches, which are able to decipher nonintuitive patterns and relationships within data sets that would otherwise remain invisible on the basis of conventional approaches.

CONCLUSIONS

In this Article, we present a machine-learning-aided platform that enables the rapid, accurate, and high-throughput screening of fly ashes by predicting a structure-based proxy for their reactivity solely on the basis of bulk chemical composition. This approach has the potential to maximize the beneficial utilization of fly ashes generated by coal power plants to replace cement in concrete, thereby generating value for what otherwise remains a costly waste (based on landfill management costs). To this end, we trained a series of neural network models based on a new data set of fly ashes sourced from the literature. On the basis of the sole knowledge of easily measurable XRF compositions, our models offer accurate predictions of the fraction of the amorphous phase in fly ashes and the composition/rigidity thereof, which play a key role in governing the reactivity of fly ashes. Importantly, our approach offers a notable improved ability to discriminate reactive from nonreactive fly ashes as compared to the conventional ASTM C618 criterion. Specifically, thanks to its increased precision and recall, our model is significantly less conservative, yet more accurate than ASTM C618 in identifying reactive fly ashes that exhibit low atomic rigidity, which, in turn, can result in additional value for power plant operators. In addition, this approach also opens new opportunities in screening fly ashes that are presently stored in impoundments, so as to reclaim ashes that had previously been unfairly classified as noncompliant by ASTM C618.

Overall, the findings presented in this study are significant in three major respects. First, the neural network models developed herein offer accurate predictions of the potential of a fly ash to replace cement in concrete based on a simple XRF measurement, which can be conducted in real-time and on-site. This is key to enable the rapid, high-throughput screening of fly ashes, so as to maximize their beneficial use. Second, given the large range of compositions of the fly ash samples comprised in the data set, the overall approach presented herein has the potential to promote the recycling of other SCMs and industrial aluminosilicate byproducts. This would require one to accumulate additional data points, identify a relevant definition or proxy for "reactivity" for the material of interest, and retrain the models introduced herein. Last, this study demonstrates that machine learning techniques can not only unlock novel perspectives for deciphering previously invisible correlations between material compositions and properties/functionalities, but also they advance our understanding of material behaviors that are governed by complex physical and chemical laws that are presently not fully clear.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssuschemeng.0c06978.

Summary of the fly ash data set; enumeration of the topological constraints in fly ashes; basics of neural networks; stratified sampling of the small data set; model optimization and data stratification; comparison of the prediction accuracy offered by different learning

https://dx.doi.org/10.1021/acssuschemeng.0c06978
ACS Sustainable Chem. Eng. 2021, 9, 2639−2650

algorithms; interpretation of the final neural network model using SHAP; and correlation between the mass fraction of CaO in the bulk material and the CAS system (PDF)
Full fly ash data set (XLSX)

AUTHOR INFORMATION

Corresponding Authors

Gaurav Sant − *Laboratory for the Chemistry of Construction Materials (LC²), 5731J Boelter Hall, Department of Civil and Environmental Engineering, Institute for Carbon Management (ICM), Department of Materials Science and Engineering, and California Nanosystems Institute, University of California, Los Angeles, California 90095, United States;* orcid.org/0000-0002-1124-5498; Email: gsant@ucla.edu

Mathieu Bauchy − *Physics of AmoRphous and Inorganic Solids Laboratory (PARISlab), 5731B Boelter Hall, Department of Civil and Environmental Engineering and Institute for Carbon Management (ICM), University of California, Los Angeles, California 90095, United States;* orcid.org/0000-0003-4600-0631; Email: bauchy@ucla.edu

Authors

Yu Song − *Physics of AmoRphous and Inorganic Solids Laboratory (PARISlab), 5731B Boelter Hall, Department of Civil and Environmental Engineering and Laboratory for the Chemistry of Construction Materials (LC²), 5731J Boelter Hall, Department of Civil and Environmental Engineering, University of California, Los Angeles, California 90095, United States;* orcid.org/0000-0001-6218-3234

Kai Yang − *Physics of AmoRphous and Inorganic Solids Laboratory (PARISlab), 5731B Boelter Hall, Department of Civil and Environmental Engineering, University of California, Los Angeles, California 90095, United States*

Jingyi Chen − *Physics of AmoRphous and Inorganic Solids Laboratory (PARISlab), 5731B Boelter Hall, Department of Civil and Environmental Engineering, University of California, Los Angeles, California 90095, United States*

Kaixin Wang − *Physics of AmoRphous and Inorganic Solids Laboratory (PARISlab), 5731B Boelter Hall, Department of Civil and Environmental Engineering, University of California, Los Angeles, California 90095, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acssuschemeng.0c06978

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge the financial support for this research by the U.S. Department of Transportation through the Federal Highway Administration (grant no. 693JJ31950021) and the U.S. National Science Foundation (DMREF: 1922167).

REFERENCES

(1) Mindess, S.; Francis, Y.; Darwin, D. *Concrete*, 2nd ed.; PCA: Upper Saddle River, NJ, 2002.

(2) Hasanbeigi, A.; Price, L.; Lin, E. Emerging Energy-Efficiency and CO2 Emission-Reduction Technologies for Cement and Concrete Production: A Technical Review. *Renewable Sustainable Energy Rev.* **2012**, *16* (8), 6220−6238.

(3) Paris, J. M.; Roessler, J. G.; Ferraro, C. C.; DeFord, H. D.; Townsend, T. G. A Review of Waste Products Utilized as Supplements to Portland Cement in Concrete. *J. Cleaner Prod.* **2016**, *121*, 1−18.

(4) Xu, G.; Shi, X. Characteristics and Applications of Fly Ash as a Sustainable Construction Material: A State-of-the-Art Review. *Resources, Conservation and Recycling* **2018**, *136*, 95−109.

(5) Kruse, K.; Andres, J.; Folliard, K.; Ferron, R.; Juenger, M.; Drimalas, T. Characterizing Fly Ash; FHWA/TX-13/0-6648-1; University of Texas at Austin, 2013.

(6) Haugsten, K. E.; Gustavson, B. Environmental Properties of Vitrified Fly Ash from Hazardous and Municipal Waste Incineration. *Waste Manage.* **2000**, *20* (2), 167−176.

(7) van Ruijven, B. J.; van Vuuren, D. P.; Boskaljon, W.; Neelis, M. L.; Saygin, D.; Patel, M. K. Long-Term Model-Based Projections of Energy Use and CO2 Emissions from the Global Steel and Cement Industries. *Resources, Conservation and Recycling* **2016**, *112*, 15−36.

(8) Prasad, B.; Sangita, K. Heavy Metal Pollution Index of Ground Water of an Abandoned Open Cast Mine Filled with Fly Ash: A Case Study. *Mine Water Environ.* **2008**, *27* (4), 265−267.

(9) American Coal Ash Association. Coal Combustion Product (CCP) Production & Use Survey Report; 2018.

(10) Schneider, M.; Romer, M.; Tschudin, M.; Bolio, H. Sustainable Cement Production—Present and Future. *Cem. Concr. Res.* **2011**, *41* (7), 642−650.

(11) Benhelal, E.; Zahedi, G.; Shamsaei, E.; Bahadori, A. Global Strategies and Potentials to Curb CO2 Emissions in Cement Industry. *J. Cleaner Prod.* **2013**, *51*, 142−161.

(12) Chancey, R. T.; Stutzman, P.; Juenger, M. C. G.; Fowler, D. W. Comprehensive Phase Characterization of Crystalline and Amorphous Phases of a Class F Fly Ash. *Cem. Concr. Res.* **2010**, *40* (1), 146−156.

(13) Oey, T.; Timmons, J.; Stutzman, P.; Bullard, J. W.; Balonis, M.; Bauchy, M.; Sant, G. An Improved Basis for Characterizing the Suitability of Fly Ash as a Cement Replacement Agent. *J. Am. Ceram. Soc.* **2017**, *100* (10), 4785−4800.

(14) ASTM C618. *Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use in Concrete*; ASTM International: West Conshohocken, PA, 2019; pp 1−5.

(15) Fox, J. M. Fly Ash Classification − Old and New Ideas, 2017; p 19.

(16) Göktepe, A. B.; Sezer, A.; Sezer, G. I.; Ramyar, K. Classification of Time-Dependent Unconfined Strength of Fly Ash Treated Clay. *Construction and Building Materials* **2008**, *22* (4), 675−683.

(17) Hower, J. C.; Mastalerz, M. An Approach toward a Combined Scheme for the Petrographic Classification of Fly Ash. *Energy Fuels* **2001**, *15* (5), 1319−1321.

(18) Suárez-Ruiz, I.; Valentim, B.; Borrego, A. G.; Bouzinos, A.; Flores, D.; Kalaitzidis, S.; Malinconico, M. L.; Marques, M.; Misz-Kennan, M.; Predeanu, G.; Montes, J. R.; Rodrigues, S.; Siavalas, G.; Wagner, N. Development of a Petrographic Classification of Fly-Ash Components from Coal Combustion and Co-Combustion. (An ICCP Classification System, Fly-Ash Working Group − Commission III.). *Int. J. Coal Geol.* **2017**, *183*, 188−203.

(19) Donatello, S.; Tyrer, M.; Cheeseman, C. R. Comparison of Test Methods to Assess Pozzolanic Activity. *Cem. Concr. Compos.* **2010**, *32* (2), 121−127.

(20) Snellings, R.; Scrivener, K. L. Rapid Screening Tests for Supplementary Cementitious Materials: Past and Future. *Mater. Struct.* **2016**, *49* (8), 3265−3279.

(21) Oey, T.; Kumar, A.; Pignatelli, I.; Yu, Y.; Neithalath, N.; Bullard, J. W.; Bauchy, M.; Sant, G. Topological Controls on the Dissolution Kinetics of Glassy Aluminosilicates. *J. Am. Ceram. Soc.* **2017**, *100* (12), 5521−5527.

(22) Traynor, B.; Mulcahy, C.; Uvegi, H.; Aytas, T.; Chanut, N.; Olivetti, E. A. Dissolution of Olivines from Steel and Copper Slags in Basic Solution. *Cem. Concr. Res.* **2020**, *133*, 106065.

(23) Snellings, R. Surface Chemistry of Calcium Aluminosilicate Glasses. *J. Am. Ceram. Soc.* **2015**, *98* (1), 303−314.

(24) Uvegi, H.; Chaunsali, P.; Traynor, B.; Olivetti, E. Reactivity of Industrial Wastes as Measured through ICP-OES: A Case Study on Siliceous Indian Biomass Ash. *J. Am. Ceram. Soc.* **2019**, *102* (12), 7678−7688.
(25) Snellings, R. Solution-Controlled Dissolution of Supplementary Cementitious Material Glasses at PH 13: The Effect of Solution Composition on Glass Dissolution Rates. *J. Am. Ceram. Soc.* **2013**, *96* (8), 2467−2475.
(26) Li, C.; Li, Y.; Sun, H.; Li, L. The Composition of Fly Ash Glass Phase and Its Dissolution Properties Applying to Geopolymeric Materials. *J. Am. Ceram. Soc.* **2011**, *94* (6), 1773−1778.
(27) Maraghechi, H.; Rajabipour, F.; Pantano, C. G.; Burgos, W. D. Effect of Calcium on Dissolution and Precipitation Reactions of Amorphous Silica at High Alkalinity. *Cem. Concr. Res.* **2016**, *87*, 1−13.
(28) Hemmings, R. T.; Berry, E. E. On the Glass in Coal Fly Ashes: Recent Advances. *MRS Online Proc. Libr. Arch.* **1987**, *113*, 1.
(29) Ward, C. R.; French, D. Determination of Glass Content and Estimation of Glass Composition in Fly Ash Using Quantitative X-Ray Diffractometry. *Fuel* **2006**, *85* (16), 2268−2277.
(30) Williams, R. P.; van Riessen, A. Determination of the Reactive Component of Fly Ashes for Geopolymer Production Using XRF and XRD. *Fuel* **2010**, *89* (12), 3683−3692.
(31) Mauro, J. C. Topological Constraint Theory of Glass. *Am. Ceram. Soc. Bull.* **2011**, *90*, 31−37.
(32) Phillips, J. C. Topology of Covalent Non-Crystalline Solids I: Short-Range Order in Chalcogenide Alloys. *J. Non-Cryst. Solids* **1979**, *34* (2), 153−181.
(33) Maxwell, J. C. L. *On the Calculation of the Equilibrium and Stiffness of Frames. London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1864**, *27* (182), 294−299.
(34) Bauchy, M. Topological Constraint Theory and Rigidity of Glasses; https://www.taylorfrancis.com/ (accessed Jun 21, 2020).
(35) Bauchy, M. Deciphering the Atomic Genome of Glasses by Topological Constraint Theory and Molecular Dynamics: A Review. *Comput. Mater. Sci.* **2019**, *159*, 95−102.
(36) Pignatelli, I.; Kumar, A.; Bauchy, M.; Sant, G. Topological Control on Silicates' Dissolution Kinetics. *Langmuir* **2016**, *32* (18), 4434−4439.
(37) Liu, H.; Zhang, T.; Anoop Krishnan, N. M.; Smedskjaer, M. M.; Ryan, J. V.; Gin, S.; Bauchy, M. Predicting the Dissolution Kinetics of Silicate Glasses by Topology-Informed Machine Learning. *npj Materials Degradation* **2019**, *3* (1), 1−12.
(38) Mascaraque, N.; Bauchy, M.; Fierro, J. L. G.; Rzoska, S. J.; Bockowski, M.; Smedskjaer, M. M. Dissolution Kinetics of Hot Compressed Oxide Glasses. *J. Phys. Chem. B* **2017**, *121* (38), 9063−9072.
(39) Mascaraque, N.; Bauchy, M.; Smedskjaer, M. M. Correlating the Network Topology of Oxide Glasses with Their Chemical Durability. *J. Phys. Chem. B* **2017**, *121* (5), 1139−1147.
(40) Oey, T.; Frederiksen, K. F.; Mascaraque, N.; Youngman, R.; Balonis, M.; Smedskjaer, M. M.; Bauchy, M.; Sant, G. The Role of the Network-Modifier's Field-Strength in the Chemical Durability of Aluminoborate Glasses. *J. Non-Cryst. Solids* **2019**, *505*, 279−285.
(41) Gin, S.; Wang, M.; Bisbrouck, N.; Taron, M.; Lu, X.; Deng, L.; Angeli, F.; Charpentier, T.; Delaye, J.-M.; Du, J.; Bauchy, M. Can a Simple Topological-Constraints-Based Model Predict the Initial Dissolution Rate of Borosilicate and Aluminosilicate Glasses? *npj Materials Degradation* **2020**, *4* (1), 1−10.
(42) Oey, T.; Plante, E. C. L.; Falzone, G.; Yang, K.; Wada, A.; Bauchy, M.; Bullard, J. W.; Sant, G. Topological Controls on Aluminosilicate Glass Dissolution: Complexities Induced in Hyper-alkaline Aqueous Environments. *J. Am. Ceram. Soc.* **2020**, *103*, 1.
(43) Oey, T.; Hsiao, Y.-H.; Callagon, E.; Wang, B.; Pignatelli, I.; Bauchy, M.; Sant, G. N. Rate Controls on Silicate Dissolution in Cementitious Environments. *RILEM Technical Letters* **2017**, *2*, 67−73.
(44) Anoop Krishnan, N. M.; Mangalathu, S.; Smedskjaer, M. M.; Tandia, A.; Burton, H.; Bauchy, M. Predicting the Dissolution Kinetics of Silicate Glasses Using Machine Learning. *J. Non-Cryst. Solids* **2018**, *487*, 37−45.
(45) Durdziński, P. T.; Dunant, C. F.; Haha, M. B.; Scrivener, K. L. A New Quantification Method Based on SEM-EDS to Assess Fly Ash Composition and Study the Reaction of Its Individual Components in Hydrating Cement Paste. *Cem. Concr. Res.* **2015**, *73*, 111−122.
(46) Scarlett, N. V. Y.; Madsen, I. C. Quantification of Phases with Partial or No Known Crystal Structures. *Powder Diffr.* **2006**, *21* (4), 278−284.
(47) Taylor-Lange, S. C.; Lamon, E. L.; Riding, K. A.; Juenger, M. C. G. Calcined Kaolinite−Bentonite Clay Blends as Supplementary Cementitious Materials. *Appl. Clay Sci.* **2015**, *108*, 84−93.
(48) Aughenbaugh, K. L.; Stutzman, P.; Juenger, M. C. G. Identifying Glass Compositions in Fly Ash. *Front. Mater.* **2016**, *3*, 1 DOI: 10.3389/fmats.2016.00001.
(49) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(50) Jablonka, K. M.; Ongari, D.; Moosavi, S. M.; Smit, B. Big-Data Science in Porous Materials: Materials Genomics and Machine Learning. *Chem. Rev.* **2020**, *120*, 8066.
(51) Liu, H.; Fu, Z.; Yang, K.; Xu, X.; Bauchy, M. Machine Learning for Glass Science and Engineering: A Review. *Journal of Non-Crystalline Solids: X* **2019**, *4*, 100036.
(52) Bagheri, A.; Nazari, A.; Sanjayan, J. The Use of Machine Learning in Boron-Based Geopolymers: Function Approximation of Compressive Strength by ANN and GP. *Measurement* **2019**, *141*, 241−249.
(53) Marks, M.; Glinicki, M. A.; Gibas, K. Prediction of the Chloride Resistance of Concrete Modified with High Calcium Fly Ash Using Machine Learning. *Materials* **2015**, *8* (12), 8714−8727.
(54) Naseri, H.; Jahanbakhsh, H.; Moghadas Nejad, F.; Golroo, A. Developing a Novel Machine Learning Method to Predict the Compressive Strength of Fly Ash Concrete in Different Ages. *AUT J. Civil Eng.* **2019**, 1.
(55) Nguyen, K. T.; Nguyen, Q. D.; Le, T. A.; Shin, J.; Lee, K. Analyzing the Compressive Strength of Green Fly Ash Based Geopolymer Concrete Using Experiment and Machine Learning Approaches. *Construction and Building Materials* **2020**, *247*, 118581.
(56) Heaton, J. *Introduction to Neural Networks with Java*; Heaton Research, Inc., 2008.
(57) Malouf, R. A Comparison of Algorithms for Maximum Entropy Parameter Estimation. *Proceedings of the 6th conference on Natural language learning - Volume 20*; COLING-02; Association for Computational Linguistics: U.S., 2002; pp 1−7.
(58) Demir-Kavuk, O.; Kamada, M.; Akutsu, T.; Knapp, E.-W. Prediction Using Step-Wise L1, L2 Regularization and Feature Selection for Small Data Sets with Large Number of Features. *BMC Bioinf.* **2011**, *12* (1), 412.
(59) Lundberg, S. M.; Lee, S. I. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*; 2017.
(60) Bauchy, M. Structural, Vibrational, and Elastic Properties of a Calcium Aluminosilicate Glass from Molecular Dynamics Simulations: The Role of the Potential. *J. Chem. Phys.* **2014**, *141* (2), No. 024507.
(61) Thomson, W.; Roth, A. E. The Shapley Value: Essays in Honor of Lloyd S. Shapley. *Economica* **1991**, *58*, 123.
(62) Varshneya, A. K. *Fundamentals of Inorganic Glasses*; Elsevier: New York, 2013.
(63) Chelberg, M. The Effect of Fly Ash Chemical Composition on Compressive Strength of Fly Ash Portland Cement Concrete; The Ohio State University: Columbus, OH, 2019.
(64) Nascimento, M. L. F.; Souza, L. A.; Ferreira, E. B.; Zanotto, E. D. Can Glass Stability Parameters Infer Glass Forming Ability? *J. Non-Cryst. Solids* **2005**, *351* (40), 3296−3308.

(65) Nascimento, M. L. F.; Ferreira, E. B.; Zanotto, E. D. Kinetics and Mechanisms of Crystal Growth and Diffusion in a Glass-Forming Liquid. *J. Chem. Phys.* **2004**, *121* (18), 8924−8928.
(66) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine Learning in Materials Informatics: Recent Applications and Prospects. *npj Computational Materials* **2017**, *3* (1), 1−13.
(67) Schmidt, J.; Marques, M. R. G.; Botti, S.; Marques, M. A. L. Recent Advances and Applications of Machine Learning in Solid-State Materials Science. *npj Computational Materials* **2019**, *5* (1), 1−36.

https://dx.doi.org/10.1021/acssuschemeng.0c06978
ACS Sustainable Chem. Eng. 2021, 9, 2639−2650