01 Mar 2023

Deep Learning to Predict the Hydration and Performance of Fly Ash-Containing Cementitious Binders

Taihao Han

Rohan Bhat

Sai Akshay Ponduru

Amit Sarkar

et. al. For a complete list of authors, see https://scholarsmine.mst.edu/ele_comeng_facwork/4606

Recommended Citation

T. Han et al., "Deep Learning to Predict the Hydration and Performance of Fly Ash-Containing Cementitious Binders," *Cement and Concrete Research*, vol. 165, article no. 107093, Elsevier, Mar 2023. The definitive version is available at https://doi.org/10.1016/j.cemconres.2023.107093

Cement and Concrete Research 165 (2023) 107093

Contents lists available at ScienceDirect

Cement and Concrete Research

journal homepage: www.elsevier.com/locate/cemconres

Deep learning to predict the hydration and performance of fly ash-containing cementitious binders

Taihao Han [a], Rohan Bhat [a], Sai Akshay Ponduru [a], Amit Sarkar [b], Jie Huang [c], Gaurav Sant [d], Hongyan Ma [e], Narayanan Neithalath [b], Aditya Kumar [a,*]

[a] Department of Materials Science and Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA
[b] School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ 85287, USA
[c] Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA
[d] Civil and Environmental Engineering, University of California, Los Angeles, Los Angeles, CA 90095, USA
[e] Civil, Architectural and Environmental Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA

ARTICLE INFO

Keywords:
Hydration
Segmentation
Deep forest
Network topology
And compressive strength

ABSTRACT

Fly ash (FA) – an industrial byproduct – is used to partially substitute Portland cement (PC) in concrete to mitigate concrete's environmental impact. Chemical composition and structure of FAs significantly impact hydration kinetics and compressive strength of concrete. Due to the substantial diversity in these physicochemical attributes of FAs, it has been challenging to develop a generic theoretical framework – and, therefore, theory-based analytical models – that could produce reliable, a priori predictions of properties of [PC + FA] binders. In recent years, machine learning (ML) – which is purely data-driven, as opposed to being derived from theorical underpinnings – has emerged as a promising tool to predict and optimize properties of complex, heterogenous materials, including the aforesaid binders. That said, there are two issues that stand in the way of widespread use of ML models: (1) ML models require thousands of data-records to *learn* input-output correlations and developing such a large, yet consistent database is impractical; and (2) ML models – while good at producing predictions – are unable to reveal the underlying correlation between composition/structure of material and its properties. This study employs a deep forest (DF) model to predict composition- and time-dependent hydration kinetics and compressive strength of [PC + FA] binders. Data dimensionality-reduction and segmentation techniques – premised on theoretical understanding of composition-structure correlations in FAs, and hydration mechanism of PC – are used to boost the DF model's prediction performance. And, finally, through inference of the intermediate and final outputs of the DF model, a simple, closed-form analytical model is developed to predict compressive strength, and reveal the correlations between mixture design and compressive strength of [PC + FA] binders.

1. Introduction

Fly ash (FA) is a byproduct of coal combustion, and, generally, treated as landfill waste. In the past decades, researchers have discovered that FA can be used to partially replace Portland cement (PC) in cementitious binders (i.e., pastes; mortars; and concrete). This is significant for reducing the carbon footprint of the PC industry. To put this in context: The manufacture of PC emits ~4–9 % of anthropogenic CO_2 emissions [1–3]; and, due to the global population growth (~82 million per year [4]), the demand for PC is expected to continually increase, thus exacerbating greenhouse gas emissions. Clean energy can be introduced to the PC manufacturing process to reduce the use of fossil fuels. However, due to technological barriers and costs/*return-on-investment* risks of upgrading manufacturing infrastructure, widespread adoption of clean energy is at least a few decades away. In 2020, clean energy made up only ~20 % [5] of utility-scale U.S. electricity generation; in developing countries, the clean energy share is even lower (i.e., <5 % [6]). Furthermore, in the PC manufacturing process, 65 % of CO_2 is released from the decomposition of $CaCO_3$; which cannot be avoided by adopting clean energy. Considering the state-of-the-art of PC manufacturing, partial substitution of PC with supplementary cementitious materials (SCMs: FA; limestone; slag; metakaolin; quartz; etc. [7–11]) is currently

* Corresponding author at: Department of Materials Science and Engineering, Missouri University of Science and Technology, B49 McNutt Hall, 1400 N Bishop, Rolla, MO 65409, USA.
E-mail address: kumarad@mst.edu (A. Kumar).

https://doi.org/10.1016/j.cemconres.2023.107093
Received 18 July 2022; Received in revised form 17 November 2022; Accepted 4 January 2023
Available online 9 January 2023

the most promising and practical solution to reduce PC's environmental impact. Among the SCMs, the use of FAs is more appealing, since it is an industrial waste, and its large-scale use would reduce the over-exploitation of geological minerals (e.g., limestone and quartz). We acknowledge that the hastening pace of coal-power plant closures – mainly due to abating costs and low carbon-footprint of renewable solar and wind energy – is fast diminishing production/reserves of FAs nationwide, thus resulting in major seasonal supply gaps, and, in turn, escalating prices. That said, based on conservative estimates of current production and historical reserves of FA, there are millions of tonnes of FA feedstocks that could be used as SCMs to formulate concrete at scale for decades to come.

Compared to other SCMs (e.g., limestone; quartz; silica fume; metakaolin; etc.), FA is arguably more complex, because of the significant batch-to-batch variations in its chemical composition; whereas, other SCMs are compositionally more consistent. FA consists of SiO_2 (abbreviate as S in cement chemistry notation), Al_2O_3 (A), CaO (C), SO_3 ($), and other metal oxides. During hydration (i.e., the reaction of PC with water [12]), FA not only provides additional surfaces (filler effect) for heterogeneous nucleation and growth [13,14] of calcium-silicate-hydrate (C—S—H; H_2O abbreviate as H), but also undergoes pozzolanic reaction with portlandite (CH; a hydration product) to form additional C—S—H [15]. Moreover, $ in FA – which is typically not found in other SCMs – can react with C_3A (a highly-reactive phase in PC) to form ettringite early on, and monosulfoaluminate at later ages. The aqueous reactivity of FAs vary over wide ranges due to disparities in their amorphous phase contents (50-to-90%$_{mass}$ [16–18]) as well as their molecular structure (i.e., topology and connectivity of the glassy network). While ASTM C618 [19] categorizes FAs into just two classes – Class C and Class F – based on their calcium content, many studies [18,20,21] have shown that this classification is not correlated with the FAs' aqueous reactivity. In recent studies, topological constraint theory (TCT) [22,23] has been used to evaluate the reactivity of amorphous materials, including FAs, based on their chemical composition and atomic structure. TCT uses the chemical composition of a dominantly amorphous material to produce a singular parameter [number of constraints (n_c)]; which represents the connectivity (or lack thereof) of the material's glassy network and can serve as a as a proxy for its reactivity. In our previous studies [24,25], we have shown that n_c is a reliable parameter to represent the reactivities of FAs.

Many studies [26–29] have replaced PC (in binders) with FA; and investigated their hydration behavior, and the resultant evolutions of the microstructure and properties. While these studies have shed light on many previously-unknown aspects of PC-FA interactions, they have been unable to fully describe the underlying correlations between composition, reaction (i.e., hydration), and properties in [PC + FA] binders. In the absence of a comprehensive theoretical framework, the route to developing reliable, theory-based analytical models remains unclear. Past studies [13,14,30–33] have shown that in PC and [PC + SCM] binders, the hydration kinetics – generally measured as time-dependent heat evolution profiles from isothermal calorimetry [34,35] – are very well correlated with properties (e.g., compressive strength). Therefore – at least in theory – if the hydration kinetics of [PC + FA] binders could be modeled, the time-dependent development of properties could be estimated, crudely if not accurately. But, on account of the compositional/structural degrees of freedom associated with FA, the complex, temporally-varying interactions between FA, PC, and hydration products (e.g., portlandite) are not well understood. This knowledge-gap – in essence – prevents the development of models that can reliably predict hydration kinetics of [PC + FA] binders. Therefore, new modeling techniques – preferably based on the data (since the theory is still unclear) – are needed to produce a priori predictions of hydration kinetics and properties of [PC + FA].

Machine learning (ML) is a promising tool for producing a priori predictions of properties of cementitious binders. Our recent studies [24,25,35–40] have shown that several ML models – including random forests, artificial neural networks, and support vector regressions models – can produce reliable predictions of heat evolution profiles and mechanical properties of cementitious materials. Although those studies have demonstrated the predictive ability of ML models, there are still a few issues that need to be addressed. (1) SCMs used in previous studies [35–37] are geological minerals (e.g., limestone; and quartz) or industrial wastes (e.g., silica fume) with consistent chemical compositions. As stated above, due to batch-to-batch variations in compositions and molecular structure, each FA affects the properties of PC binders with a distinct footprint. The prediction performance of ML models on such *highly heterogeneous* systems (i.e., [PC + FA] binders), especially in terms of predicting time-dependent hydration kinetics, has not been investigated. (2) Our previous papers [24,25] used TCT along with ML to predict the properties of FA-based alkali-activated binders. However, it remains unclear if TCT can serve as a useful data-distillation tool to extract critical information on [PC + FA] binders to enhance the learning-efficacy of ML models. Furthermore, FAs used in these prior studies were mostly devoid of $; whereas FAs in this study have up to ~8 % $ content. As mentioned before, $ in FAs can exert substantial effects on both fresh and hardened properties of PC binders. This study introduces an additional input variable to account for $ content; which TCT – in the way it has been formulated and used in prior studies [18,24,25,41] – is unable to account for. (3) Han et al. [36] have shown that Fourier transform can reduce the degrees of freedom of the training database; thus enhancing the learning-efficacy of ML models. That said, the Fourier transform is a purely mathematical tool used, primarily, to process nonlinear sensor signals; and, thus, remains completely divorced from properties of the material. In materials-based research, it would arguably be useful if the mathematical tools were based on, or derived from, materials-based theories. Therefore, this study develops a hydration theory-guided approach (i.e., segmentation) to simplify the database of hydration kinetics of [PC + FA] binders. The segmentation technique allows ML models to produce reliable predictions by learning from a small dataset.

In this study, a deep forest (DF) model is utilized to predict time- and composition-dependent hydration kinetics and compressive strength in relation to the mixture design of [PC + FA] binders and physicochemical properties of 10 FAs. The DF model is coupled with the segmentation technique to enhance the prediction performance. Finally, through inference of the intermediate and final outputs of the DF model, a simple, closed-form analytical model is developed to predict compressive strength, and reveal the correlations between mixture design and compressive strength of [PC + FA] binders.

2. Experimental methods

All precursors used in this study were obtained from commercial suppliers. Two PCs (i.e., PC-1 and PC-2) were used as base material, and 10 FAs (Boral Resources) were used as supplementary cementitious materials to substitute PCs at different replacement levels. Based on ASTM C150 [42], PC-1 and PC-2 were classified in Type I/II. All precursors were received in a powder form.

According to ASTM D4326-13 [43], the chemical compositions of PCs and FAs were quantified using the x-ray fluorescence (XRF; Oxford X-supreme 8000). Here, XRF – a simple, easy-to-use characterization tool—is used to evaluate FAs; because the goal of this study to demonstrate that bulk oxide composition of FAs can be used—as the sole and primary input—to not only estimate its chemo-structural attributes (i.e., *number of constraints*), but also to predict its influence of hydration kinetics and mechanical properties. The phase compositions (Table 1) of two PCs were calculated by the modified Bogue method [44] based on their XRF compositions. The oxide compositions of FAs are shown in Table 2. The particle size distributions of the PCs and FAs were measured by a static light scattering particle size analyzer (Microtrac S3500). All particles were assumed to be spherical shape [45]. During the measurement, the PCs and FAs were suspended in isopropanol and deionized

Table 1
Phase composition and specific surface area (SSA) of the PCs used in this study.

Cement type	C3S	C2S	C3A	C4AF	Gypsum	d50	SSA
	$\%_{mass}$	$\%_{mass}$	$\%_{mass}$	$\%_{mass}$	$\%_{mass}$	μm	$cm^2.g^{-1}$
PC-1	71.98	6.05	5.48	11.07	5.41	40.57	1726
PC-2	52.61	24.88	5.61	10.21	6.67	38.21	1814

water, respectively. The suspended powders were treated with ultrasonic pulses to avoid agglomeration. The specific surface areas (SSA) of the PCs and FAs were estimated from their particle size distributions [13,46]. We assumed that all particles had a spherical morphology, limited surface roughness, and zero internal porosity. The density of the PCs is 3.15 $g.cm^{-3}$, and SSAs are shown in Table 1. The densities (provided by the supplier) and SSAs of the FAs are shown in Table 2.

The heat evolution profiles of [PC + FA] binders were measured by the I-Cal 8000 isothermal calorimetry for 72 h at 20 °C. The binders were prepared by mixing solid with deionized water in a constant liquid-to-solid ratio (0.4). Isothermal calorimetry was used to measure heat flow rate and cumulative heat from 102 unique mixture designs (i.e., 50 [PC-1 + FA] binders, 50 [PC-2 + FA] binders, and 2 plain PC binders). The replacement levels of FAs varied from 10-to-50 % by mass with 10 % intervals.

The compressive strength was tested in accordance with the ASTM C109 [47] standard. The compressive strength for each binder was calculated as the average of measurements of 3-to-4 specimens. The testing specimens included all binders for which calorimetry experiments were also conducted; except for the [PC + FA-10] binder. All specimens were cured at >98 % relative humidity and 22 °C for 3, 7, and 28 days. The data obtained from calorimetry and compressive strength test are used to train and validate the DF model.

3. Modeling methods

3.1. Network topology

A singular, numerical parameter (*number of constraints*) obtained from the network topology of FAs is employed to replace all chemical descriptors (i.e., XRF composition) of FAs in the ML database; which not only reduces the complexity of the database (by substantially reducing the number of input variables associated with FAs), but also produces a singular parameter encapsulating information regarding the molecular structure and aqueous reactivity of the FAs. This section demonstrates a concise description of *number of constraints* derived from the TCT; further details can be found in our recent publications [24,25].

FAs consist of three components: S; C; and A. Other minor components are ignored within the TCT framework. The normalized chemical composition of FA is $C_xA_yS_{1-x-y}$, where x and y are the molar fraction. Two chemical constraints exist in FAs: angular bond-bending (BB) and radial bonding-stretching (BS) constraints [41,48,49]. Si/Al tetrahedrons provide 4 BS constraints and 5 BB constraints, O atoms connected with Si/Al tetrahedrons have 1 BB constraint, and O atoms connected with Ca atoms carry 1 BS constraint [41,49–52]. Ca atoms do not have constraints. In TCT, the chemical structure of the FAs can be divided into three groups based on their chemical compositions: depolymerized regime ($y - x \leq -\frac{2}{3}$); partially depolymerized regime ($-\frac{2}{3} \leq y - x \leq 0$); and fully polymerized regime ($0 \leq y - x$). In the fully depolymerized regime, calcium (Ca) becomes the dominant compositional descriptor (vis-à-vis other atoms present within the structure). This regime, typically, features the isolation of the Si and Al tetrahedrons from each other because of interactions with Ca and O to form abundant non-bridging oxygens (NBOs). The NBOs reduce long-range order structures, thereby enhancing aqueous reactivity of the material upon contact with alkaline solutions (e.g., concrete pore-solution). In the partially depolymerized regime, Si is the dominant component dictating network connectivity (or lack thereof). This regime consists of both bridging oxygen (BOs) atoms and NBOs. Due to the presence of BOs, the Si and Al tetrahedrons form rigid structures, resulting in an increase in crystalline content and a reduction of reactivity. In the case of the fully polymerized regime, Al atoms are the most dominant. To form Al tetrahedrons, Ca needs to compensate charge from Al atom. Due to insufficient free Ca atoms, few non-bridging oxygens are present in the structure. This leads to a highly crystalline structure, leading to the lowest reactivity. Among the three regimes, fully depolymerized regime has the highest reactivity, whereas fully polymerized regime has the lowest reactivity.

The methods used to calculate the *number of constraints* (n_c) in each regime are shown in Eqs. (1a), (1b) and (1c). In this study, the FAs are in partially depolymerized and fully polymerized regimes. The *number of constraints* for each FA is itemized in Table 2.

$$n_c = \frac{11 + y - 10x}{3 - 2x + 2y} \text{ (Fully depolymerized)} \tag{1a}$$

$$n_c = \frac{11 + 10y - 10x}{3 - 2x + 2y} \text{ (Partially depolymerized)} \tag{1b}$$

$$n_c = \frac{11 + 13y - 13x}{3 - 2x + 2y} \text{ (Fully polymerized)} \tag{1c}$$

3.2. Deep forest (DF) model

DF model is a modified version of the classification-and-regression trees (CARTs) model converged with the bagging algorithm [53,54].

Table 2
Oxide composition, density, specific surface area (SSA), and number of constraints (n_c) of FAs used in this study.

Oxide composition	FA-1	FA-2	FA-3	FA-4	FA-5	FA-6	FA-7	FA-8	FA-9	FA-10
	($\%_{mass}$)	($\%_{mass}$)	($\%_{mass}$)	($\%_{mass}$)	($\%_{mass}$)	($\%_{mass}$)	($\%_{mass}$)	($\%_{mass}$)	($\%_{mass}$)	($\%_{mass}$)
SiO_2	51.9	43.6	22.3	77.1	56.8	57.7	30.3	37.7	33.6	53.8
$Al2O_3$	20.4	19.0	10.2	11.1	19.2	19.9	16.8	20.9	18.1	15.4
CaO	15.6	4.6	45.6	2.4	6.2	10.7	32.6	24.7	30.4	14.0
SO_3	1.5	4.0	7.4	0.0	1.3	0.4	2.8	1.2	2.0	1.0
Na_2O	0.0	0.0	0.9	1.3	0.7	0.1	1.1	1.3	1.2	2.4
MgO	1.7	0.0	2.6	0.0	1.6	2.2	7.9	4.3	5.2	3.3
Fe_2O_3	5.5	23.5	8.7	2.4	10.2	6.0	5.6	6.3	6.4	6.5
Others	3.4	5.3	2.3	5.7	4.0	2.9	3.0	3.6	3.1	3.5
Density ($g.cm^{-3}$)	2.32	2.51	2.77	2.50	2.42	2.59	2.90	2.76	2.86	2.64
d50 (μm)	34.77	35.56	16.62	8.69	24.79	12.57	11.55	7.45	9.11	18.43
SSA ($cm^2.g^{-1}$)	2889	2022	3350	5657	3301	4419	4565	6378	5259	3505
n_c (unitless)	3.6128	3.8519	2.8831	3.7510	3.7810	3.6740	3.2870	3.4791	3.3546	3.5953

At the training stage, the DF model randomly extracts data-records from the training dataset to construct many small datasets named bootstraps. Then, the bootstraps are used as roots to grow independent CARTs; here, each bootstrap only grows one tree. The DF model can grow hundreds to thousands of CARTs. Binary split is the mechanism of growing CARTs. At each node, the tree randomly chooses a subset of input variables, and ranks them based on their costs for the split. The optimal split (global minimum cost) is estimated; and the tree is allowed to grow until the homogeneity of terminal leaves cannot be further improved (i.e., convergence is reached). The DF model allows all trees to grow as deep as they can without smoothing and pruning. At the testing stage, the DF model averages parallel outputs from trees to calculate the final output. The DF model involves two-stage randomization [54,55]: data selection for bootstrap and variable selection at each split. Owing to this feature, CARTs in the forest decorrelate with one another, leading to a minimization of bias and variance errors [56,57]. In this study, to optimize the hyperparameters of the DF model, the 10-fold cross-validation (CV) method [39,58] integrated with the grid-search method [36,39] is applied to the model. This allows the DF model to archive the ideal prediction performance.

3.3. Database collection

The heat evolution profiles of [PC + FA] binders were obtained from isothermal calorimetry (see description of the method in Section 2). The parent database consists of time-dependent heat flow rates from 111 distinct [PC + FA] binders. It should be noted that calorimetry experiments of select [PC + FA] binders were repeated to verify the repeatability and reproducibility of the method; as well as to estimate the standard deviation (<0.005 mW/g_{cement}) associated with each measurement. But, in the parent database, results from such duplicate measurements were not included. The testing dataset comprises of 10 randomly-selected binders from the parent database; and the remaining binders are consolidated into the training dataset. It is worth pointing out, here, that the binders included in the testing dataset were not included in the training dataset. This was done to ensure that the prediction performance of the DF model could be evaluated in "blind-tests". The DF model establishes the underlying input-output relationships from the training dataset. The testing dataset is then used to evaluate the prediction-performance of the DF model. Both the training and testing datasets comprise binders formulated using 2 PCs and 10 FAs. The input variables are: C_3S content ($\%_{mass}$); C_2S content ($\%_{mass}$); C_3A content ($\%_{mass}$); C_4AF content ($\%_{mass}$); gypsum content ($\%_{mass}$); FA content ($\%_{mass}$); number of constraints (unitless); $ content in FA ($\%_{mass\ in\ FA}$); SSA of PC ($cm^2.\ g^{-1}$); SSA of FA ($cm^2.\ g^{-1}$); and time (hour). The output is the heat flow rate ($mW.g_{cem}^{-1}$) from 0-to-72 h, with a 1-h time interval between successive steps. Statistical parameters pertaining to the training and testing datasets are itemized in Tables S1 and S2.

The compressive strengths of [PC + FA] binders were obtained from tests described in Section 2. The database including compressive strengths of 92 unique [PC + FA] binders (i.e., 2 plain PC and 90 PCs replaced by FAs). For each binder, at least 4 samples were evaluated to estimate the standard deviation (i.e., <4.2 MPa); and the average of these measurements were used in the parent database. The database is randomly split into training and testing dataset. The training dataset includes 75 % of the parent database, and the remaining data-records are included in the testing dataset. The input variables include: C_3S content ($\%_{mass}$); C_2S content ($\%_{mass}$); C_3A content ($\%_{mass}$); C_4AF content ($\%_{mass}$); gypsum content ($\%_{mass}$); FA content ($\%_{mass}$); number of constraints (unitless); $ content in FA ($\%_{mass\ in\ FA}$); SSA of PC ($cm^2.\ g^{-1}$); SSA of FA ($cm^2.\ g^{-1}$); and age (day). The output is compressive strength (MPa) at 3, 7, and 28 days. Statistical parameters pertaining to the training and testing datasets are itemized in Table S3.

Five statistical parameters are calculated to assess the accuracy of predictions of the DF model on the testing datasets: mean absolute percentage error (*MAPE*); mean absolute error (*MAE*); Pearson correlation coefficient (*R*); root mean squared error (*RMSE*); and coefficient of determination (R^2). Equations for statistical parameters were shown in prior research [39,59,60].

4. Results and discussion

4.1. Isothermal calorimetry of [PC + FA] binders

Heat evolution profiles of [PC + FA] binders were measured by isothermal calorimetry. Fig. 1 demonstrates the representative heat evolution profiles of PC replaced by 40 % FAs. The 40 % replacement level was chosen because it adequately exhibits both the physical and chemical influences of FAs on the hydration kinetics of PC. Fig. 1a and c show the influences of fully polymerized FAs on heat flow rate and cumulative heat, and Fig. 1b and d show the influences of partially depolymerized FAs.

In Fig. 1a, it is clear that PC-1 replaced by FA-4 has the shortest induction period and the highest heat flow peak. This is because FA-4 does not contain $ (which would be manifested as delayed hydration peak due to retardation of C_3S hydration); but has the highest S content (77.1$\%_{mass}$), which enhances—albeit, to a limited degree—the formation of pozzolanic C—S—H. Furthermore, the additional surfaces provided by FA-4 accelerate the nucleation and growth of C—S—H formed from the hydration of the alite phase in PC. Owing to these factors, the [PC-1 + FA-4] binder releases more heat than plain PC paste (shown in Fig. 1c). The hydration peak of [PC-1 + FA-2] binder appears late and broader compared to other binders. This broadening of the hydration peak (and, thus, greater area under the differential heat curve) is also why the [PC-1 + FA-2] binder releases slightly more heat than the [PC-1 + FA-4] binder (shown in Fig. 1c). These differences in the hydration profile of [PC-1 + FA-2] binder with respect to [PC-1 + FA-4] binder arise primarily because of the relatively higher $ content (4 %) in FA-2. To be clear: Both [PC-1 + FA-2] and [PC-1 + FA-4] binders comprise FAs with high pozzolanic potential; but FA-2 comprises more $ to react with C_3A (in PC) to form ettringite and monosulfoaluminate, thus releasing more heat, while also suppressing the hydration of C_3S (and, thus, resulting in delayed hydration peak). Furthermore, the formation of C—S—H (via hydration of alite and belite, or via pozzolanic reaction) releases less (cumulative) heat compared to the formation of ettringite and monosulfoaluminate (via hydration of calcium aluminate + sulfate phases) [12,61]. Therefore, FAs with higher $ content, typically, result in greater cumulative heat release at later ages.

The fully polymerized FAs have rigid molecular structures; which, in turn, makes then less reactive, and less susceptible to dissolving and releasing ions into the contiguous aqueous solutions. Upon comparison of heat flow rates between PC binders replaced by FA-2 (Fig. 1a) and FA-9 (Fig. 1b), it is found that FA-9 (a partially depolymerized FA) suppresses the hydration of PC for a longer time, although it has approximately half of the $ content as FA-2. The partially depolymerized FAs – including FA-9 –have more amorphous content, and are typically more reactive, than fully polymerized FAs [18,49]. Therefore, a larger abundance of ions (including $, which could suppress C_3S hydration) are expected to be released from partially depolymerized FAs compared to fully polymerized ones. In Fig. 1b, the PC replaced by FA-3 shows the highest hydration peak. This is because FA-3 has the lowest *number of constraints* (2.88), which entails it has the highest amorphous content among all FAs. This implies that FA-3 is the most reactive among all FAs; and can release ions at a faster rate upon contact with the solution. However, the S content in FA-3 is low; as such, its ability to partake in pozzolanic reaction is expected to be limited (despite faster release of silicate ions), especially at later ages; thereby, resulting in moderate cumulative heat release (Fig. 1d). In fact, in general, partially depolymerized FAs have lower S content than fully polymerized FAs; and, therefore, have limited potential to partake in pozzolanic reactions (shown in Fig. 1c and d). Another important point to note is that the aluminate ions from FAs – especially in FAs with high Al_2O_3 content –

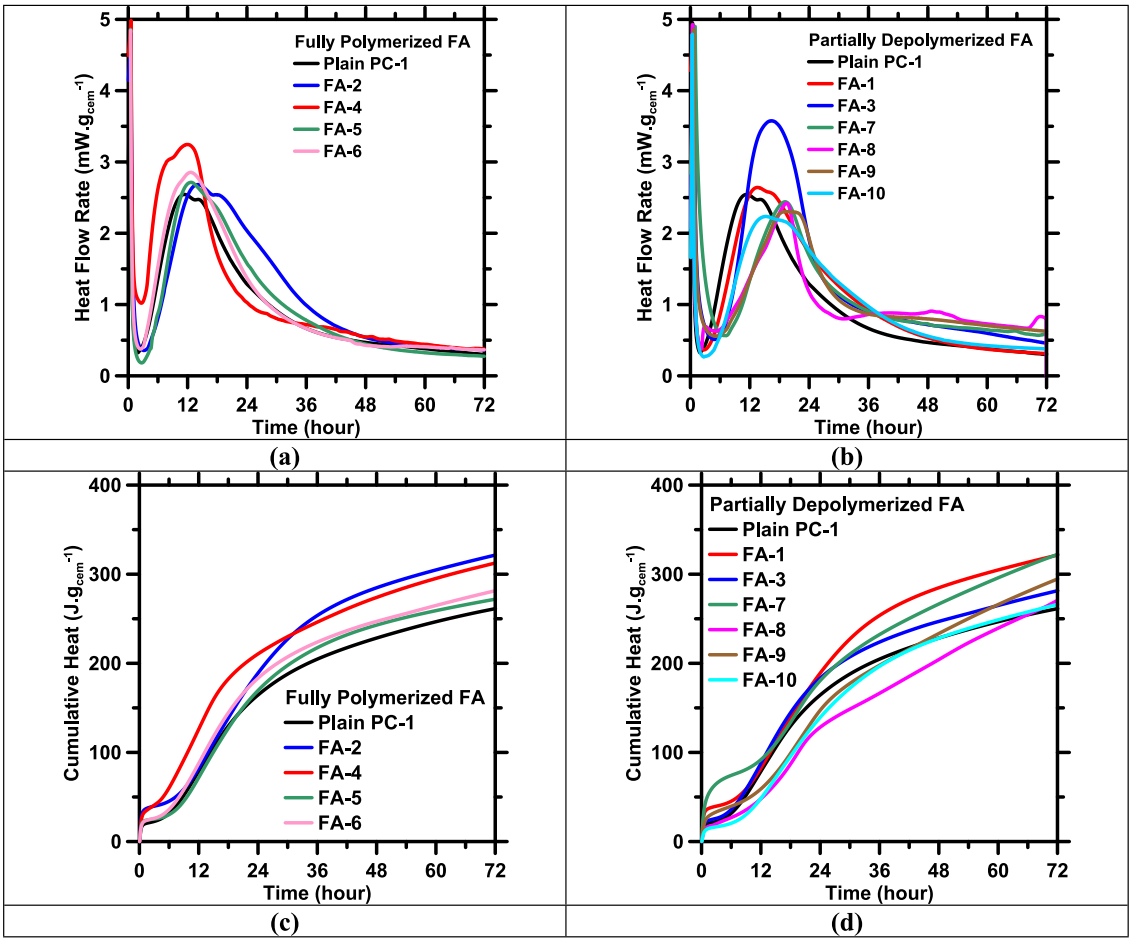

Fig. 1. Hydration kinetic profiles of [PC-1 + 40 % FA] binders, that is, (a) heat flow rate of PC-1 replaced by fully polymerized FA; (b) heat flow rate of PC-1 replaced by partially depolymerized FA; (c) cumulative heat of PC replaced by fully polymerized FA; and (d) cumulative heat of PC replaced by partially depolymerized FA. The FA type is shown in the legends. All measurements were recorded for initial 72 h of hydration.

could suppress the growth of C—S—H [15]. This can be verified by a sharper heat flow rate decline after the hydration peaks (Fig. 1b) in the binder containing FA-8 as compared to the binder formulated using FA-3.

Fig. 2 shows heat evolution profiles of PC-2 replaced by FA-4 at different replacement levels; so as to demonstrate the influence of FA content on hydration kinetics. Here, [PC-2 + FA-4] binder has been chosen as the representative binder because the composition of FA-4 is that of typical, commercial FA (rich in aluminosilicate content, and devoid of $). With increasing FA-4 content, the filler effect becomes more pronounced, which manifests as heat flow rate progressively shifting to the left, slope for hydration peak becoming steeper, and

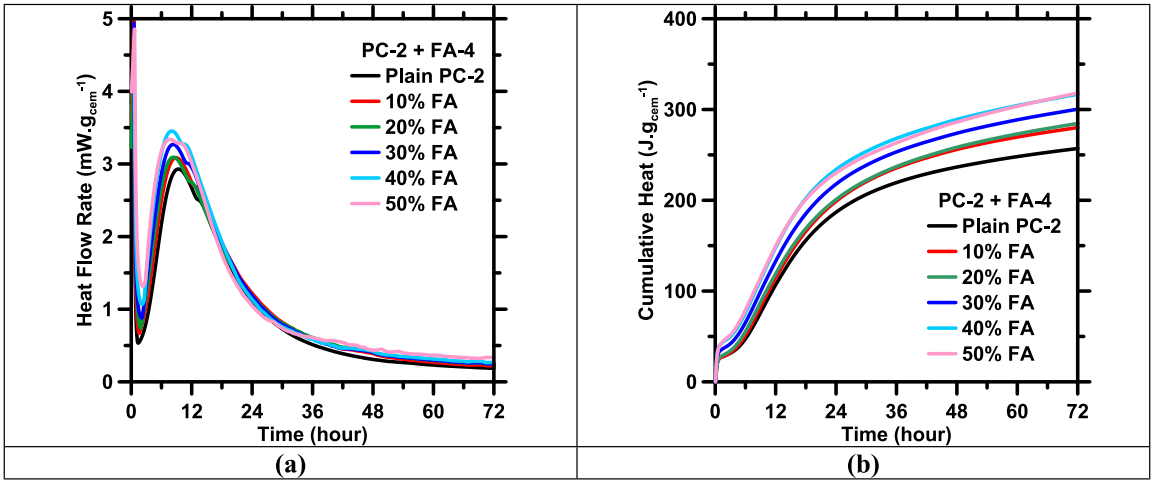

Fig. 2. Hydration kinetic profiles of [PC-2 + FA-4] binders, that is, (a) heat flow rate and (b) of PC replaced by different amount of FA. The replacement level of FA is shown in the legends. All measurements were recorded for initial 72 h of hydration.

induction period getting shortened. High replacement level of FA-4 provides additional surfaces for heterogeneous nucleation of C—S—H. The increment of hydration peak intensity, and cumulative heat also indicates that the high replacement level enhances the kinetics of precipitation of C—S—H.

4.2. Predictions of hydration kinetics

The DF model is used to predict heat flow rate profiles of 10 randomly-selected [PC + FA] binders for a 72-h hydration period. Fig. 3 shows predicted heat flow rate profiles of representative [PC + FA] binders as produced by the standalone DF model against experimental measurements. In order to visually compare the predicted and measured values, the entire heat flow rate spectrums are included. The statistical parameters pertaining to the prediction performance of 72 h hydration period on the testing dataset are itemized in Table 3.

As can be seen in Table 3, predictions of heat flow rate as produced by DF model show reasonable accuracy, with R^2 being 0.89 and *RMSE* being 0.24 mW.g_{cem}^{-1}. In Fig. 3, it is worth pointing out that the predictions at early ages, especially at the hydration peak, show lower accuracy than at later ages. At early ages, the heat flow rate changes rapidly; sharply declining from very high to very low values, and subsequently rising quickly until the peak. Each FA – in relation to its content and physicochemical properties – influences heat flow rate in a unique way. Because those factors act coherently and simultaneously within a very short span of time (a few hours), the DF is unable to fully capture the cause-effect correlations during its training; which essentially diminishes its ability to predict early-age hydration kinetics of [PC + FA] binders in a high-fidelity manner.

Table 3
Statistical parameters pertaining the prediction performance of DF and [DF + Segmentation] models on 72-h heat flow rate.

ML model	R	R^2	MAE	MAPE	RMSE
	Unitless	*Unitless*	mW. g_{cem}^{-1}	*%*	mW. g_{cem}^{-1}
DF	0.9476	0.8981	0.1471	13.27	0.2463
DF + Segmentation	0.9871	0.9743	0.0701	7.43	0.1197

It is important to further finetune the DF model to produce reliable predictions of heat flow rate, especially for a small volume database. Towards this end, the hydration theory-guided segmentation technique is developed to reduce the complexity of the database and enhance the prediction accuracy. The segmentation technique is based on the hypothesis that, in the same segment, the hydration behavior of all [PC + FA] binders should be driven by the same mechanisms, and therefore should manifest as similar kinetic (heat evolution) profiles. Here, the PC hydration can be divided into four stages [62]: initial dissolution period; induction period; acceleration period; and deceleration period. The dissolution period is a very short stage, where the PC particles come into contact with water, and release a large amount of heat. After ions are released from the PC particles, the dissolution rate of the PC particles significantly reduces (i.e., 17 orders of magnitude [43,44]) while the

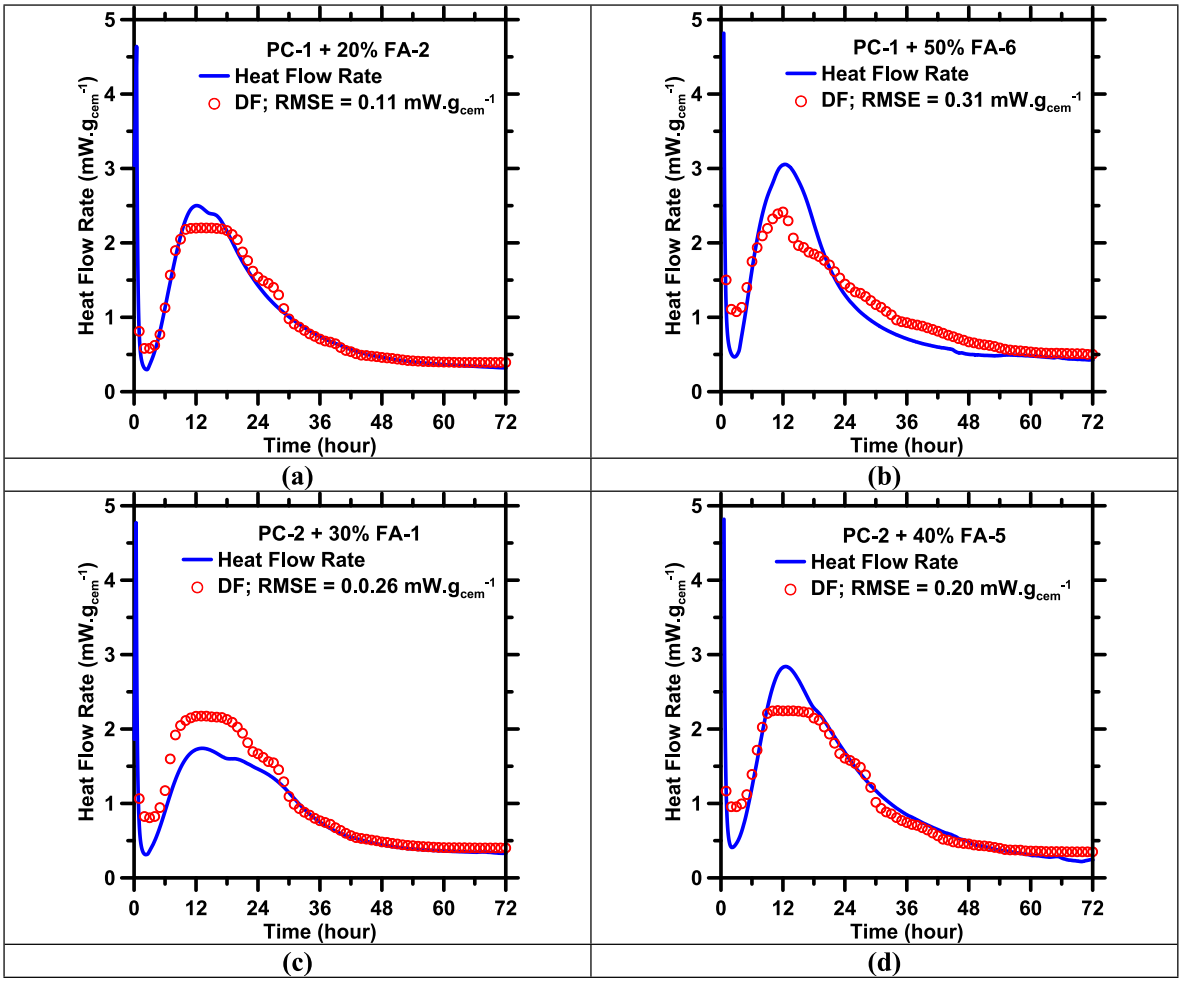

Fig. 3. The standalone DF model predictions of heat flow rate of representative [PC + FA] binders: (a) PC-1 + 20 % FA-2; (b) PC-1 + 50 % FA-6; (c) PC-2 + 30 % FA-1; and (d) PC-2 + 40 % FA-5 compared against experimental measurements in the testing database. Root mean square errors (*RMSE*) of predictions are shown in legends.

solution is still undersaturated. The slow dissolution period is the induction period. During this period, nuclei of hydrates form on the surfaces of PC particles. When the solution is supersaturated with respect to the hydrates (portlandite and C—S—H), the rate of nucleation and growth accelerates. The heat evolution profile of PC shows a large exothermic heat release peak for the nucleation and growth. The post-peak period represents decelerating hydration kinetics. In the deceleration period, several factors reduce the hydration rate [62]: (1) Only large anhydrous particles contribute to hydration; (2) Lack of space and exponentially-declining supersaturation of C—S—H constrain the rate of growth of C—S—H nuclei; and (3) Insufficient water in the system limits hydration rates of anhydrous phases present within PC particles.

In accordance with the above description of stages of PC hydration, the heat flow rate can be divided into three segments with respect to time. The first segment includes the initial dissolution and induction periods. In this segment, high heat release is observed in the first hour, and the heat flow rate for the remaining duration is low. The second segment is the acceleration period, where the hydration kinetics is the fastest. One or two hydration peaks – depending on the chemical compositions of FAs and PC – appear in this segment. The third segment is the deceleration period, where the heat flow rate progressively diminishes after the peak, due to the continued decline in the kinetics of hydration. We developed a MATLAB algorithm to find the optimal thresholds for the segmentation. To obtain these thresholds, the algorithm obtains the second derivates of heat flow rate profiles, and determines the inflection points. The hours corresponding to the majority of inflection points (across all binders) are determined as segmentation thresholds. Fig. 4 shows the thresholds for the different segments in representative binders. The induction period for all [PC + FA] binders end before 5 h. All hydration peaks appear no later than 22 h, and the deceleration periods start after 24 h. Accordingly, the three segments are 0-to-4, 5-to-24, and 25-to-72 h.

The DF model integrated with segmentation technique is developed in this study. When the [DF + Segmentation] model processes the training dataset, the segmentation algorithm separates the training dataset into three sub-datasets with respect to the abovementioned times. Then, three parallel DF models are independently trained with three sub-datasets to find the input-output correlations for each segment. When the model is used to produce predictions of hydration kinetics of binders in the testing dataset, three parallel DF models predict the heat flow rate of [PC + FA] binders with respect to their segments. Later, the segmentation algorithm combines the outcomes from DF models (employed for the different segments) to produce the entire heat evaluation profile from the time of mixing to 72 h.

Fig. 5 shows predicted heat flow rate profiles of representative [PC + FA] binders as produced by the [DF + Segmentation] model compared against experimental measurements. In order to visually compare the predicted and measured values, the entire heat flow rate spectrums are included. The statistical parameters pertaining to the prediction performance of 72 h hydration period on the testing dataset are itemized in Table 3. To better compare the prediction performance between the standalone DF model and [DF + Segmentation] model, Fig. 6 demonstrates the prediction accuracy (evaluated by five statistical parameters) of both models for every hour.

As can be seen in Table 3, predictions of heat flow rate as produced by DF model show high accuracy, with R^2 being 0.97 and *RMSE* being 0.11 mW.g_{cem}^{-1}. The errors – in terms of *MAE*, *MAPE*, and *RMSE* – of [DF + Segmentation] model are halved when compared with the standalone DF model. In Figs. 5 and 6, it can be observed that the [DF + Segmentation] model is able to capture cause-effect correlations and produce reliable trends during the acceleration period, especially the hydration peaks. The segmentation technique reduces the degree of freedom of the database by splitting the complex, highly non-monotonous heat evolution profiles into segments with more-or-less similar profiles. Following the segmentation of the training database, the DF model can learn input-output correlations (hydration behavior) in each segment; and, later, consolidate them into global *composition-hydration kinetics* correlations. It must also be noted that the use of TCT to reduce the number of input variables (i.e., content of each oxide) associated with each FA into a singular parameter also contributes to the prediction performance of the models.

In the context of modeling heat evolution profiles of cementitious systems, it is essential to capture the heat evolution behaviors at peaks; this is because these peaks represent a critical juncture of PC hydration, and are indicative of the time when the PC binders rapidly gain strength due to the massive formation of hydrates [12,32,62,63,64]. Due to the importance of hydration peaks, previous studies [13,34,65] have utilized the ability to capture hydration peaks as an exclusive indicator to assess the accuracy of kinetic models. Since the standalone DF model is unable to produce reliable predictions at hydration peaks, it is justified to say that the [DF + Segmentation] model is a more suitable tool to predict heat evaluation profiles of complex PC binders. For some binders (Fig. 5b), the [DF + Segmentation] still has some issues in predicting peaks in a high-fidelity manner. The database used to train the DF model is relatively small, containing only 101 binders. A larger and more diverse database could significantly enhance the prediction accuracy of the model.

4.3. Predictions of compressive strength

The database of compressive strength was split into training (75 % of parent database) and testing (25 % of parent database) datasets. The predictions of compressive strength of [PC + FA] binders as produced by the DF model against measurements, including both training and testing

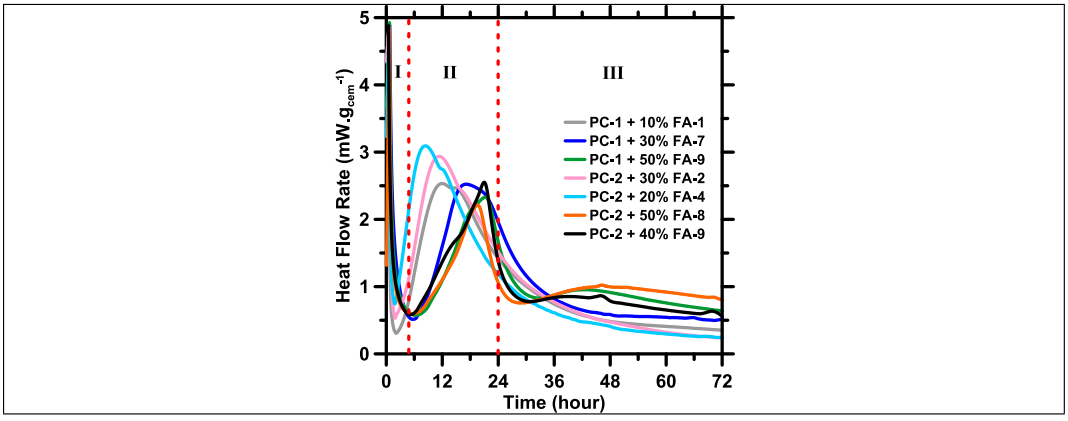

Fig. 4. Heat flow rate profiles of randomly selected [PC + FA] binders. Based on their kinetic behaviors, the profiles are divided into three segments: I. Initial and induction periods; II. Acceleration period; and III. Deceleration period.

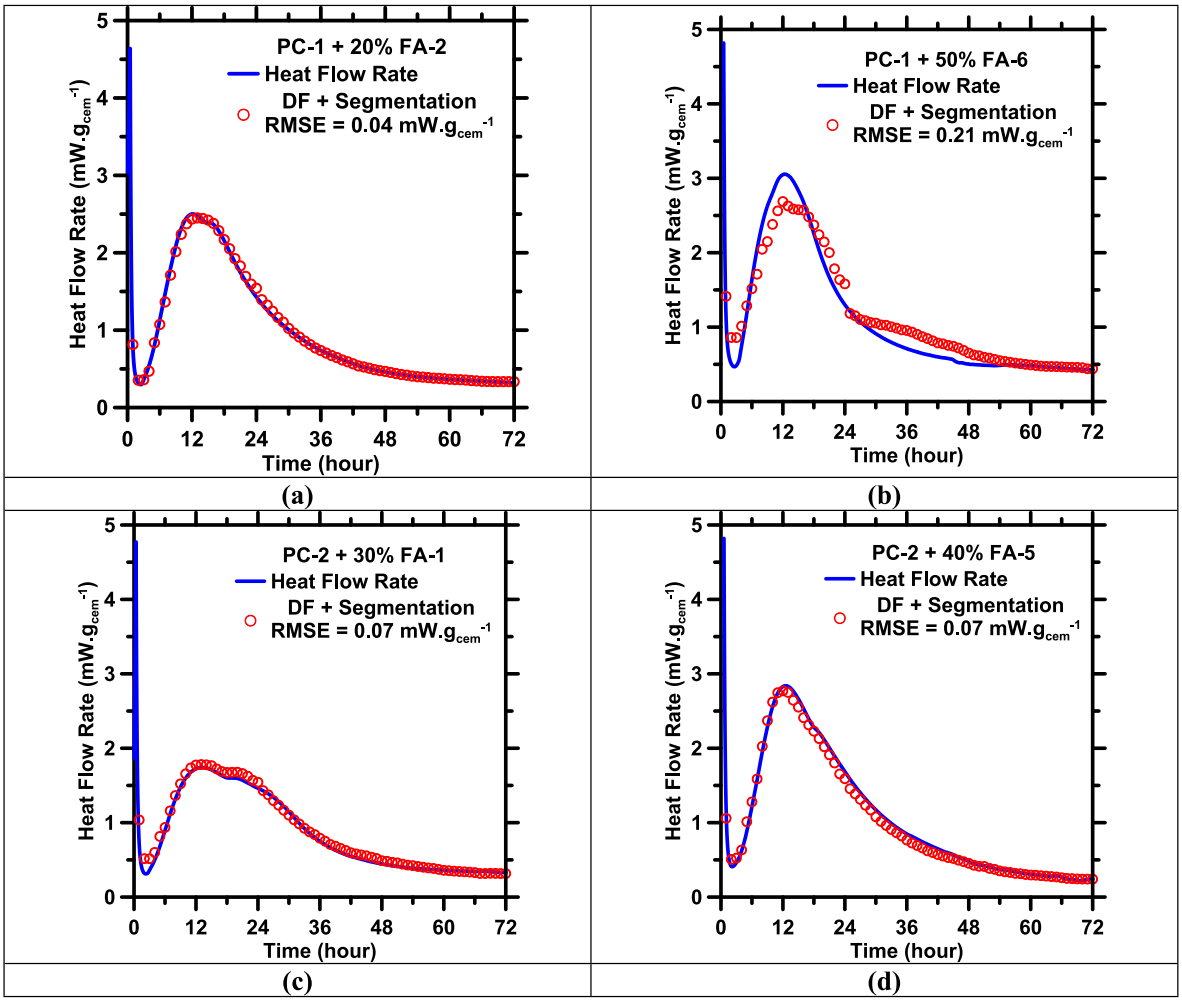

Fig. 5. The [DF + Segmentation] model predictions of heat flow rate of representative [PC+ FA] binders: (a) PC-1 + 20 % FA-2; (b) PC-1 + 50 % FA-6; (c) PC-2 + 30 % FA-1; and (d) PC-2 + 40 % FA-5 compared against experimental measurements in the testing database. Root mean square errors (*RMSE*) of predictions are shown in legends.

datasets, are shown in Fig. 7. Predictions for the training dataset are more accurate than the testing dataset; this is because the model was trained with an emphasis on reducing overfitting. Statistical parameters evaluating the prediction performance of the DF model are itemized in Table 4.

As can be seen in Fig. 7 and Table 4, predictions of compressive strength against the testing dataset are reliable, with R^2 being and *RMSE* being 5.22 MPa. The typical measurement error of compressive strength is 5 MPa [66], where the prediction error is close to the experimental error. The prediction accuracy of the DF model can be attributed to its intrinsic structure, especially of the decision-trees in the ensemble. The forest contains hundreds of independent trees, and each of them produces one output, thereby minimizing error associated with variance. Due to the two-stage randomization [54,55], and the unpruned structure of all the trees, each tree can grow with unique structures, which eliminates bias (overfitting). Lastly, the use of the TCT significantly minimizes the number of input variables pertaining to FAs; which, in turn, improves the learning-efficiency of the model during its training, and ultimately bolsters its prediction performance.

One major issue of ML models is that they cannot quantitatively reveal correlations between inputs and output. Consequently, researchers cannot decipher the complex mixture design-compressive strength relationships of [PC + FA] binders through ML models. To expand the use of ML, herein, a simple, closed-form analytical model is developed to predict compressive strength of [PC + FA] binders through inference of the intermediate and final outputs of the DF model.

The most important step in developing an analytical model is to carefully select input variables. The exclusion of insignificant input variables reduces the complexity of model without lowering prediction accuracy. The DF model can quantify the impacts of the input variables on the output based on the Gini score [40,54,59,67,68]. Fig. 8 shows the impact (importance) of input variables on the compressive strength of [PC + FA] binders. This outcome is in good agreement with findings from prior literature. Age is the most important variable because compressive strength monotonically increases with the time of hydration [69–71]. C_3S and C_3A are significant input variables because their hydration products are the main phases that contribute to microstructural evolution, solid-to-solid phase connectivity, and therefore strength. Surprisingly, C_4AF exhibits more importance than C_3S. Due to the large content of C_3S in PC, all binders gain high compressive strength. The variation of compressive strength with respect to C_3S content is smaller than the variation with respect to C_4AF. FA replacement level is the fourth important input variable. Fu et al. [72] have shown that compressive strength of PC is inversely correlated with aggregate content. Although FAs can react with PC through pozzolanic reaction, the major fraction of FAs do not partake in chemical reaction (akin to aggregates). Due to the low concentration, $ in FA show insignificant influence on the compressive strength.

The importance of input variables can be used as a guideline to select features for the analytical model. Previous studies [73,74] have found

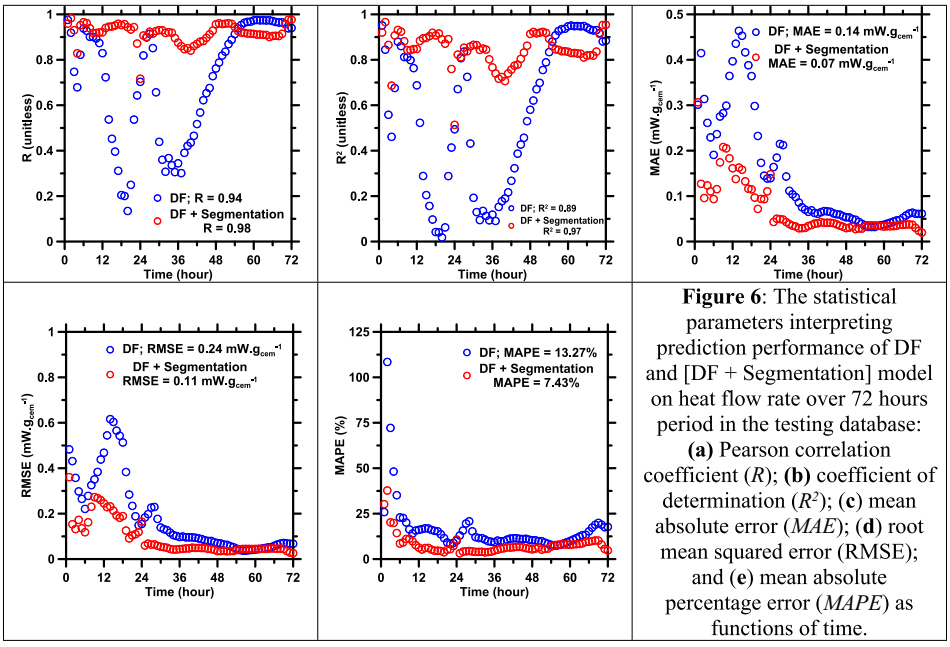

Figure 6: The statistical parameters interpreting prediction performance of DF and [DF + Segmentation] model on heat flow rate over 72 hours period in the testing database: **(a)** Pearson correlation coefficient (*R*); **(b)** coefficient of determination (R^2); **(c)** mean absolute error (*MAE*); **(d)** root mean squared error (RMSE); and **(e)** mean absolute percentage error (*MAPE*) as functions of time.

Fig. 6. The statistical parameters interpreting prediction performance of DF and [DF + Segmentation] model on heat flow rate over 72 h period in the testing database: (a) Pearson correlation coefficient (*R*); (b) coefficient of determination (R^2); (c) mean absolute error (*MAE*); (d) root mean squared error (RMSE); and (e) mean absolute percentage error (*MAPE*) as functions of time.

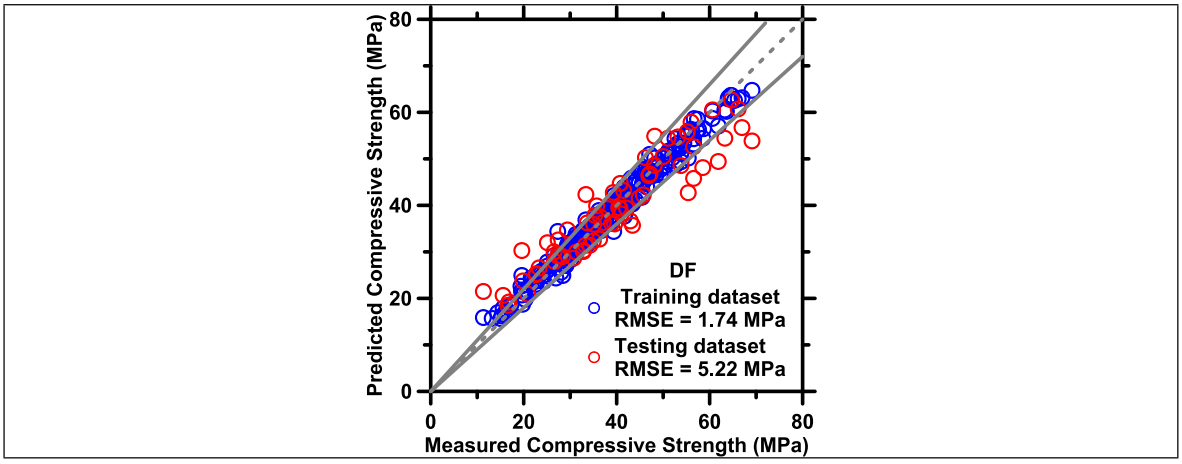

Fig. 7. The predictions of compressive strength of [PC + FA] binders as produced by the DF model against the testing dataset. Root mean square error (*RMSE*) is shown in the legend. The dashed line is the ideal prediction. The solid lines are ±10 % error bounds.

Table 4
The statistical parameters pertaining to the prediction performance of the DF and analytical models on [PC+ FA] binders' compressive strength.

ML model	R	R^2	MAE	MAPE	RMSE
	Unitless	*Unitless*	MPa	%	MPa
DF	0.9368	0.8775	3.939	11.34	5.221
Analytical model	0.9031	0.8156	4.051	11.6	5.322

that the age-compressive strength relationship is exponential for cementitious materials. The correlation between content of binder and compressive strength exhibits linearity [75,76]. All chemical phases of PC are included in the analytical model. Owing to the influences of C_4AF and C_3S, they are assigned greater weight (i.e., higher exponent) in the analytical model. Considering the pozzolanic reaction of FAs, the *number of constraints* is added as an input to the analytical model. $ in FA has a similar role as gypsum, and, thus, it is combined with the gypsum content in the analytical model. SSAs of PC and FA are excluded from the analytical model because they show less influence on compressive strength. A combination of weighted linear, polynomial, and exponential functions is utilized to develop the general form of the analytical model, which is shown in Eq. (2). Here, *CS* is compressive strength (MPa); C_i is coefficient; X_i is chemical content ($\%_{mass}$) of [PC + FA] binder; x_{so3} is mass fraction of $ content in FA (unitless); *T* is age (days); and n_c is number of constraints (unitless). The coefficients and constant in the equation are optimized through Nelder-Mead multi-dimensional simplex algorithm [77,78] and a nonlinear, gradient-descent scheme [40,59,79,80]. The optimized values of coefficients and constant are itemized in Table 5.

$$CS = C_1X_{C3S}^2 + C_2X_{C2S} + C_3X_{C3A} + C_4X_{C4AF}^2 + C_5\left(X_{gypsum} + X_{FA}x_{SO3}\right) + C_6X_{FA}N_c + C_7exp\left(\frac{C_8}{T}\right) + C_9 \quad (2)$$

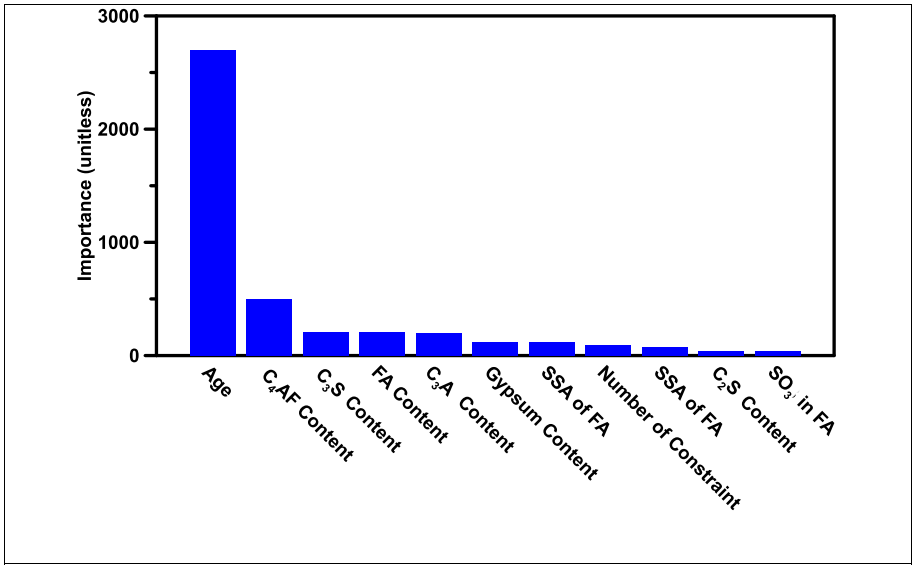

Fig. 8. The influence of input variables on compressive strength of [PC + FA] binders ranked by the DF model. Input variables are ranked in a descending order based on their influences.

Fig. 9 shows the predicted compressive strength as produced by the analytical model against experimental measurements. Statistical parameters evaluating the prediction performance of the analytical model are summarized in Table 4. The analytical model predicts the compressive strength of [PC + FA] binders in a high-fidelity manner. The *RMSE* obtained from predictions of the analytical model is 5.32 MPa, which—expectedly—is slightly larger than predictions from the DF model. It should be noted that the redundant variables were identified and excluded from the analytical model; and, then, coefficients and constant were rigorously optimized. As a result, the analytical model uses fewer input variables without sacrificing much accuracy. The major advantage of the analytical model is to uncover the underlying correlations between input variables and compressive strength. Importantly, this analytical model accounts for physicochemical properties and $ content of FAs. This helps researchers to understand the role of FAs on the compressive strength of PC in a quantitative manner.

5. Conclusions

FA is a type of industry waste used to partially substitute PC in concrete to mitigate concrete's environmental impact. FAs – depending on chemical composition; molecular structure; and replacement level – show disparate influences on hydration kinetics and compressive strength of PC binders. Due to the complexities associated with FAs' highly-heterogenous composition, and our piecemeal understanding of material theories, researchers have been unable to develop reliable analytical models to predict properties of [PC + FA] binders.

Table 5
Coefficients and constant for the analytical model to achieve optimal predictions of compressive strength of [PC + FA] binders.

C_1	−0.0031
C_2	−0.7886
C_3	5.286
C_4	−0.3974
C_5	−1.9971
C_6	−0.3068
C_7	7.2686
C_8	−9.9688
C_9	109.6953

In this study, we harnessed the power of the deep forest (DF) model to produce a priori, reliable predictions of hydration kinetics and compressive strength of [PC + FA] binders. The DF model can autonomously learn cause-effect relationships from a training dataset, and then capitalize on such knowledge to produce predictions on new data domains. Additionally, this study developed a material theory-guided segmentation technique to enhance the prediction accuracy of hydration kinetics. The segmentation method classified the heat flow rate profiles into three regions based on the time-dependent hydration behavior of cement. Next, after the DF model discovered the input-compressive strength correlations, key outcomes (i.e., quantitative ranking of influence of each variable on compressive strength) were used as guidelines for developing a simple, closed-form analytical model. Results from the analytical model demonstrated that it was a reliable tool to produce a priori predictions of compressive strength by using fewer input variables than ML. Researchers could use correlations shown in the analytical model to investigate mechanisms that the origins of property-development in [PC + FA] binders.

In summary, the DF model with the segmentation technique demonstrates the possibility of using a small volume database to produce reliable predictions of complex cementitious materials. The prediction accuracy can be improved by training the DF model with a larger database. One limitation of this study is that the segmentation is a mathematical method, which provides little-to-no insight into aspects of cement chemistry that could affect hydration kinetics. In an upcoming study, the ML model will be integrated with a phase boundary nucleation-and-growth based kinetic model to produce hydration kinetics for cementitious binders.

CRediT authorship contribution statement

Aditya Kumar, Hongyan Ma, Jie Huang, Gaurav Sant, and Narayanan Neithalath: Conceptualization, analyses, writing – review and editing, funding acquisition, supervision, project administration.

Amit Sarkar, Rohan Bhat, and Sai Akshay Ponduru: Production of experimental data, data analyses and inference, writing – original draft, and preparing graphs.

Taihao Han: Development, training, and testing of machine learning model.

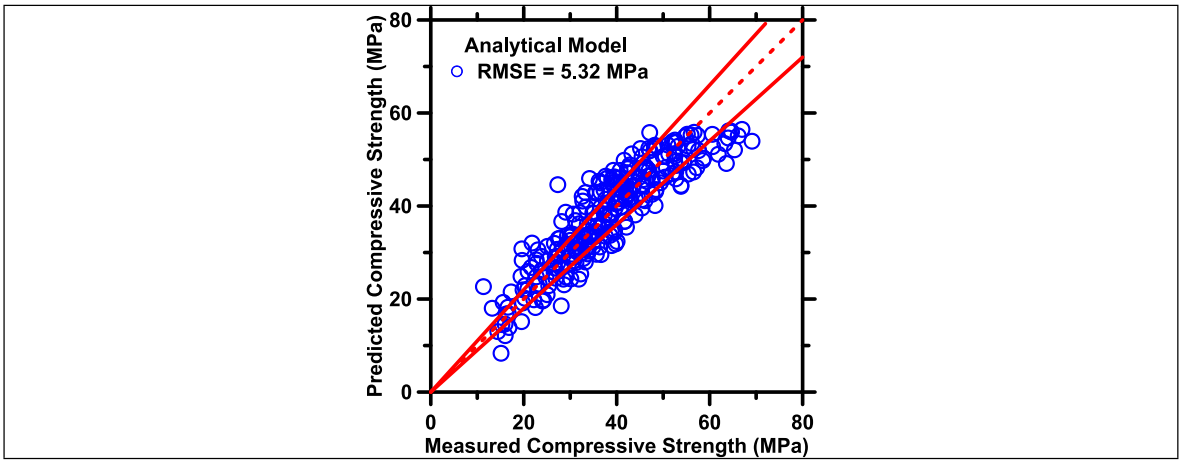

Fig. 9. The predictions of compressive strength of [PC + FA] binders as produced by the analytical model against the testing dataset. Coefficient of determination (R^2) is shown in the legend. The dashed line is the ideal predictions. The solid lines are ±10 % error bounds.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors acknowledge financial support from the Kummer Institute (Missouri S&T) Ignition Grant; the Leonard Wood Institute (LWI: W911NF-07-2-0062); the National Science Foundation (NSF-CMMI: 1932690; NSF-DMR: 2034856; and NSF-DMR: 2228782); and the Federal Highway Administration (Award no: 693JJ31950021).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cemconres.2023.107093.

References

[1] A. Dowling, J. O'Dwyer, C.C. Adley, Lime in the limelight, J. Clean. Prod. 92 (2015) 13–22, https://doi.org/10.1016/j.jclepro.2014.12.047.
[2] F. Schorcht, Best Available techniques (bat) reference document for the production of cement, lime and magnesium oxide: industrial emissions directive. https://ec.europa.eu/jrc/en/publication/reference-reports/best-available-techniques-bat-reference-document-production-cement-lime-and-magnesium-oxide, 2013.
[3] M.M. Miller, Lime, United States Geological Survey, 2012 Minerals Yearbook, 2013.
[4] United Nations, Growing at a slower pace, world population is expected to reach 9.7 billion in 2050 and could peak at nearly 11 billion around 2100. https://www.un.org/development/desa/en/news/population/world-population-prospects-2019.html, 2019.
[5] Renewable Energy, Center for Climate and Energy Solutions. (n.d.). https://www.c2es.org/content/renewable-energy.
[6] H. Hu, N. Xie, D. Fang, X. Zhang, The role of renewable energy consumption and commercial services trade in carbon dioxide reduction: evidence from 25 developing countries, Appl. Energy 211 (2018) 1229–1244, https://doi.org/10.1016/j.apenergy.2017.12.019.
[7] J.J. Biernacki, J.W. Bullard, G. Sant, K. Brown, F.P. Glasser, S. Jones, T. Ley, R. Livingston, L. Nicoleau, J. Olek, Others, cements in the 21st century: challenges, perspectives, and opportunities, J. Am. Ceram. Soc. 100 (2017) 2746–2773, https://doi.org/10.1111/jace.14948.
[8] M.C.G. Juenger, R. Snellings, S.A. Bernal, Supplementary cementitious materials: New sources, characterization, and performance insights, Cem. Concr. Res. 122 (2019) 257–273, https://doi.org/10.1016/j.cemconres.2019.05.008.
[9] M. Juenger, J.L. Provis, J. Elsen, W. Matthes, R.D. Hooton, J. Duchesne, L. Courard, H. He, F. Michel, R. Snellings, et al., Supplementary cementitious materials for concrete: characterization needs, in: MRS Proceedings, Cambridge University Press, 2012 imrc12-1488.
[10] M.C.G. Juenger, R. Siddique, Recent advances in understanding the role of supplementary cementitious materials in concrete, Cem. Concr. Res. 78 (Part A) (2015) 71–80, https://doi.org/10.1016/j.cemconres.2015.03.018.
[11] M. Johari, J.J. Brooks, S. Kabir, P. Rivard, Influence of supplementary cementitious materials on engineering properties of high strength concrete, Constr. Build. Mater. 25 (2011) 2639–2648, https://doi.org/10.1016/j.conbuildmat.2010.12.013.
[12] H.F. Taylor, Cement Chemistry, Thomas Telford, 1997.
[13] T. Oey, A. Kumar, J.W. Bullard, N. Neithalath, G. Sant, The filler effect: the influence of filler content and surface area on cementitious reaction rates, J. Am. Ceram. Soc. 96 (2013) 1978–1990, https://doi.org/10.1111/jace.12264.
[14] A. Kumar, T. Oey, S. Kim, D. Thomas, S. Badran, J. Li, F. Fernandes, N. Neithalath, G. Sant, Simple methods to estimate the influence of limestone fillers on reaction and property evolution in cementitious materials, Cem. Concr. Compos. 42 (2013) 20–29, https://doi.org/10.1016/j.cemconcomp.2013.05.002.
[15] J. Lapeyre, A. Kumar, Influence of pozzolanic additives on hydration mechanisms of tricalcium silicate, J. Am. Ceram. Soc. 101 (2018) 3557–3574, https://doi.org/10.1111/jace.15518.
[16] S.V. Vassilev, C.G. Vassileva, Methods for characterization of composition of fly ashes from coal-fired power stations: a critical overview, Energy Fuel 19 (2005) 1084–1098, https://doi.org/10.1021/ef049694d.
[17] C.R. Ward, D. French, Determination of glass content and estimation of glass composition in fly ash using quantitative X-ray diffractometry, Fuel 85 (2006) 2268–2277, https://doi.org/10.1016/j.fuel.2005.12.026.
[18] Y. Song, K. Yang, J. Chen, K. Wang, G. Sant, M. Bauchy, Machine learning enables rapid screening of reactive fly ashes based on their network topology, ACS Sustain. Chem. Eng. 9 (2021) 2639–2650, https://doi.org/10.1021/acssuschemeng.0c06978.
[19] C09 Committee, C618-15 Standard Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use in Concrete, ASTM International, West Conshohocken, PA, 2015, https://doi.org/10.1520/C0618-15.
[20] T. Oey, J. Timmons, P. Stutzman, J.W. Bullard, M. Balonis, M. Bauchy, G. Sant, An improved basis for characterizing the suitability of fly ash as a cement replacement agent, J. Am. Ceram. Soc. 100 (2017) 4785–4800, https://doi.org/10.1111/jace.14974.
[21] J.C. Hower, M. Mastalerz, An approach toward a combined scheme for the petrographic classification of fly ash, Energy Fuel 15 (2001) 1319–1321, https://doi.org/10.1021/ef010146z.
[22] J.C. Mauro, Topological constraint theory of glass, Am. Ceram. Soc. Bull. 90 (2011) 31–37.
[23] J.C. Phillips, Topology of covalent non-crystalline solids I: short-range order in chalcogenide alloys, J. Non-Cryst. Solids 34 (1979) 153–181, https://doi.org/10.1016/0022-3093(79)90033-4.
[24] R. Bhat, T. Han, S. Akshay Ponduru, A. Reka, J. Huang, G. Sant, A. Kumar, Predicting compressive strength of alkali-activated systems based on the network topology and phase assemblages using tree-structure computing algorithms, Constr. Build. Mater. 336 (2022), 127557, https://doi.org/10.1016/j.conbuildmat.2022.127557.
[25] T. Han, E. Gomaa, A. Gheni, J. Huang, M. ElGawady, A. Kumar, Machine learning enabled closed-form models to predict strength of alkali-activated systems, J. Am. Ceram. Soc. 105 (2022) 4414–4425, https://doi.org/10.1111/jace.18399.
[26] A. Oner, S. Akyuz, R. Yildiz, An experimental study on strength development of concrete containing fly ash and optimum usage of fly ash in concrete, Cem. Concr. Res. 35 (2005) 1165–1171, https://doi.org/10.1016/j.cemconres.2004.09.031.
[27] K. De Weerdt, M.B. Haha, G. Le Saout, K.O. Kjellsen, H. Justnes, B. Lothenbach, Hydration mechanisms of ternary Portland cements containing limestone powder and fly ash, Cem. Concr. Res. 41 (2011) 279–291, https://doi.org/10.1016/j.cemconres.2010.11.014.

[28] A.L.A. Fraay, J.M. Bijen, Y.M. de Haan, The reaction of fly ash in concrete a critical examination, Cem. Concr. Res. 19 (1989) 235–246, https://doi.org/10.1016/0008-8846(89)90088-4.

[29] Kou Shi Cong, Poon chi sun, Chan Dixon, influence of fly ash as cement replacement on the properties of recycled aggregate concrete, J. Mater. Civ. Eng. 19 (2007) 709–717, https://doi.org/10.1061/(ASCE)0899-1561(2007)19:9(709).

[30] D.P. Bentz, C.F. Ferraris, S.Z. Jones, D. Lootens, F. Zunino, Limestone and silica powder replacements for cement: early-age performance, Cem. Concr. Compos. 78 (2017) 43–56, https://doi.org/10.1016/j.cemconcomp.2017.01.001.

[31] W. Meng, A. Kumar, K. Khayat, Effect of silica fume and slump-retaining PCE on the development of properties of Portland cement paste, Cem. Concr. Compos. 99 (2019) 181–190, https://doi.org/10.1016/j.cemconcomp.2019.03.021.

[32] I. Mehdipour, A. Kumar, K.H. Khayat, Rheology, hydration, and strength evolution of interground limestone cement containing PCE dispersant and high volume supplementary cementitious materials, Mater. Des. 127 (2017) 54–66, https://doi.org/10.1016/j.matdes.2017.04.061.

[33] A. Kumar, T. Oey, G.P. Falla, R. Henkensiefken, N. Neithalath, G. Sant, A comparison of intergrinding and blending limestone on reaction and strength evolution in cementitious materials, Constr. Build. Mater. 43 (2013) 428–435, https://doi.org/10.1016/j.conbuildmat.2013.02.032.

[34] A.M. Ley-Hernandez, J. Lapeyre, R. Cook, A. Kumar, D. Feys, Elucidating the effect of water-to-cement ratio on the hydration mechanisms of cement, ACS Omega 3 (2018) 5092–5105, https://doi.org/10.1021/acsomega.8b00097.

[35] J. Lapeyre, T. Han, B. Wiles, H. Ma, J. Huang, G. Sant, A. Kumar, Machine learning enables prompt prediction of hydration kinetics of multicomponent cementitious systems, Sci. Rep. 11 (2021) 3922, https://doi.org/10.1038/s41598-021-83582-6.

[36] T. Han, S.A. Ponduru, R. Cook, J. Huang, G. Sant, A. Kumar, A deep learning approach to design and discover sustainable cementitious binders: strategies to learn from small databases and develop closed-form analytical models, Front. Mater. 8 (2022), 796476, https://doi.org/10.3389/fmats.2021.796476.

[37] R. Cook, T. Han, A. Childers, C. Ryckman, K. Khayat, H. Ma, J. Huang, A. Kumar, Machine learning for high-fidelity prediction of cement hydration kinetics in blended systems, Mater. Des. 208 (2021), 109920, https://doi.org/10.1016/j.matdes.2021.109920.

[38] E. Gomaa, T. Han, M. ElGawady, J. Huang, A. Kumar, Machine learning to predict properties of fresh and hardened alkali-activated concrete, Cem. Concr. Compos. 115 (2021), 103863, https://doi.org/10.1016/j.cemconcomp.2020.103863.

[39] R. Cook, J. Lapeyre, H. Ma, A. Kumar, Prediction of compressive strength of concrete: a critical comparison of performance of a hybrid machine learning model with standalone models, ASCE J. Mater. Civ. Eng. 31 (2019) 04019255, https://doi.org/10.1061/(ASCE)MT.1943-5533.0002902.

[40] T. Han, A. Siddique, K. Khayat, J. Huang, A. Kumar, An ensemble machine learning approach for prediction and optimization of modulus of elasticity of recycled aggregate concrete, Constr. Build. Mater. 244 (2020), 118271, https://doi.org/10.1016/j.conbuildmat.2020.118271.

[41] K. Yang, Y. Hu, Z. Li, N.M.A. Krishnan, M.M. Smedskjaer, C.G. Hoover, J.C. Mauro, G. Sant, M. Bauchy, Analytical model of the network topology and rigidity of calcium aluminosilicate glasses, J. Am. Ceram. Soc. 104 (2021) 3947–3962, https://doi.org/10.1111/jace.17781.

[42] C01 Committee, C150-07 Specification for Portland Cement, ASTM International, West Conshohocken, PA, 2012, https://doi.org/10.1520/C0150-07.

[43] D05 Committee, D4326-13 Standard Test Method for Major and Minor Elements in Coal and Coke Ash by X-Ray Fluorescence, ASTM International, West Conshohocken, PA, 2013, https://doi.org/10.1520/D4326-13.

[44] H.F.W. Taylor, Modification of the Bogue calculation, Adv. Cem. Res. 2 (1989) 73–77, https://doi.org/10.1680/adcr.1989.2.6.73.

[45] E.J. Garboczi, J.W. Bullard, Shape analysis of a reference cement, Cem. Concr. Res. 34 (2004) 1933–1937, https://doi.org/10.1016/j.cemconres.2004.01.006.

[46] R. Cook, H. Ma, A. Kumar, Influence of size-classified and slightly soluble mineral additives on hydration of tricalcium silicate, J. Am. Ceram. Soc. 103 (2020) 2764–2779, https://doi.org/10.1111/jace.16936.

[47] ASTM International, C109/C109M-20b standard test method for compressive strength of hydraulic cement mortars (Using 2-in. Or [50-mm] cube specimens), ASTMInternational (2020), https://doi.org/10.1520/C0109_C0109M-20B.

[48] M. Bauchy, Deciphering the atomic genome of glasses by topological constraint theory and molecular dynamics: a review, Comput. Mater. Sci. 159 (2019) 95–102, https://doi.org/10.1016/j.commatsci.2018.12.004.

[49] T. Oey, A. Kumar, I. Pignatelli, Y. Yu, N. Neithalath, J.W. Bullard, M. Bauchy, G. Sant, Topological controls on the dissolution kinetics of glassy aluminosilicates, J. Am. Ceram. Soc. 100 (2017) 5521–5527, https://doi.org/10.1111/jace.15122.

[50] M. Bauchy, M.J. Abdolhosseini Qomi, C. Bichara, F.-J. Ulm, R.J.-M. Pellenq, Nanoscale structure of cement: viewpoint of rigidity theory, J. Phys. Chem. C 118 (2014) 12485–12493, https://doi.org/10.1021/jp502550z.

[51] M. Bauchy, M. Micoulaut, Atomic scale foundation of temperature-dependent bonding constraints in network glasses and liquids, J. Non-Cryst. Solids 357 (2011) 2530–2537, https://doi.org/10.1016/j.jnoncrysol.2011.03.017.

[52] T. Oey, K.F. Frederiksen, N. Mascaraque, R. Youngman, M. Balonis, M. M. Smedskjaer, M. Bauchy, G. Sant, The role of the network-modifier's field-strength in the chemical durability of aluminoborate glasses, J. Non-Cryst. Solids 505 (2019) 279–285, https://doi.org/10.1016/j.jnoncrysol.2018.11.019.

[53] L. Breiman, Bagging predictors, Mach. Learn. 24 (1996) 123–140, https://doi.org/10.1007/BF00058655.

[54] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[55] A. Liaw, M. Wiener, Classification and Regression by RandomForest, 2001.

[56] Gäs.L. Biau, Consistency of random forests and other averaging classifiers, J. Machine Learning Research 9 (2008) 2015–2033.

[57] X. Chen, H. Ishwaran, Random forests for genomic data analysis, Genomics 99 (2012) 323–329, https://doi.org/10.1016/j.ygeno.2012.04.003.

[58] C. Schaffer, Selecting a classification method by cross-validation, Mach. Learn. 13 (1993) 135–143, https://doi.org/10.1007/BF00993106.

[59] T. Han, N. Stone-Weiss, J. Huang, A. Goel, A. Kumar, Machine learning as a tool to design glasses with controlled dissolution for application in healthcare industry, Acta Biomater. 107 (2020) 286–298, https://doi.org/10.1016/j.actbio.2020.02.037.

[60] R. Cai, T. Han, W. Liao, J. Huang, D. Li, A. Kumar, H. Ma, Prediction of surface chloride concentration of marine concrete using ensemble machine learning, Cem. Concr. Res. 136 (2020), 106164, https://doi.org/10.1016/j.cemconres.2020.106164.

[61] W. Kurdowski, Cement and Concrete Chemistry, Springer, Dordrecht, New York, 2014.

[62] J.W. Bullard, H.M. Jennings, R.A. Livingston, A. Nonat, G.W. Scherer, J. S. Schweitzer, K.L. Scrivener, J.J. Thomas, Mechanisms of cement hydration, Cem. Concr. Res. 41 (2011) 1208–1223, https://doi.org/10.1016/j.cemconres.2010.09.011.

[63] E.M. Gartner, H.M. Jennings, Thermodynamics of calcium silicate hydrates and their solutions, J. Am. Ceram. Soc. 70 (1987) 743–749, https://doi.org/10.1111/j.1151-2916.1987.tb04874.x.

[64] S. Garrault, A. Nonat, Hydrated layer formation on tricalcium and dicalcium silicate surfaces: experimental study and numerical simulations, Langmuir 17 (2001) 8131–8138, https://doi.org/10.1021/la011201z.

[65] A. Kumar, S. Bishnoi, K.L. Scrivener, Modelling early age hydration kinetics of alite, Cem. Concr. Res. 42 (2012) 903–918, https://doi.org/10.1016/j.cemconres.2012.03.003.

[66] E. Nawy (Ed.), Concrete Construction Engineering Handbook, CRC Press, 2008, https://doi.org/10.1201/9781420007657.

[67] X. Xu, T. Han, J. Huang, A.A. Kruger, A. Kumar, A. Goel, Machine learning enabled models to predict sulfur solubility in nuclear waste glasses, ACS Appl. Mater. Interfaces (2021), https://doi.org/10.1021/acsami.1c10359.

[68] B.H. Menze, B.M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, F. A. Hamprecht, A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data, BMC Bioinformatics 10 (2009) 213, https://doi.org/10.1186/1471-2105-10-213.

[69] P. Termkhajornkit, Q.H. Vu, R. Barbarulo, S. Daronnat, G. Chanvillard, Dependence of compressive strength on phase assemblage in cement pastes: beyond gel–space ratio — experimental evidence and micromechanical modeling, Cem. Concr. Res. 56 (2014) 1–11, https://doi.org/10.1016/j.cemconres.2013.10.007.

[70] S. Aggoun, M. Cheikh-Zouaoui, N. Chikh, R. Duval, Effect of some admixtures on the setting time and strength evolution of cement pastes at early ages, Constr. Build. Mater. 22 (2008) 106–110, https://doi.org/10.1016/j.conbuildmat.2006.05.043.

[71] G.F. Kheder, A.M.A. Gabban, S.M. Abid, Mathematical model for the prediction of cement compressive strength at the ages of 7 and 28 days within 24 hours, Mater. Struct. 36 (2003) 693, https://doi.org/10.1007/BF02479504.

[72] T.C. Fu, W. Yeih, J.J. Chang, R. Huang, The influence of aggregate size and binder material on the properties of pervious concrete, Adv. Mater. Sci. Eng. 2014 (2014), e963971, https://doi.org/10.1155/2014/963971.

[73] W. Abbass, M.I. Khan, S. Mourad, Experimentation and predictive models for properties of concrete added with active and inactive SiO2 fillers, Materials 12 (2019) 299, https://doi.org/10.3390/ma12020299.

[74] S. Gavela, N. Nikoloutsopoulos, G. Papadakos, D. Passa, A. Sotiropoulou, Multifactorial experimental analysis of concrete compressive strength as a function of time and water-to-cement ratio, in: Procedia Structural Integrity 10, 2018, pp. 135–140, https://doi.org/10.1016/j.prostr.2018.09.020.

[75] S.M. AL-Shukaili, Q.B. Alias, I. Latif, Multi variable empirical formula for SFRC compressive, tensile, and shear strength, in: International Conference on Civil, Disaster Management and Environmental Sciences, 2017.

[76] X.-Y. Wang, Y. Luan, Modeling of hydration, strength development, and optimum combinations of cement-slag-limestone ternary concrete, Int. J. Concr. Struct. Mater. 12 (2018) 12, https://doi.org/10.1186/s40069-018-0241-z.

[77] J.A. Nelder, R. Mead, A simplex method for function minimization, Comput. J. 7 (1965) 308–313.

[78] K.I.M. McKinnon, Convergence of the nelder-Mead simplex method to a nonstationary point, SIAM J. Optim. 9 (1998) 148–158, https://doi.org/10.1137/S1052623496303482.

[79] J.H. Friedman, Stochastic gradient boosting, in: Computational Statistics & Data Analysis 38, 2002, pp. 367–378, https://doi.org/10.1016/S0167-9473(01)00065-2.

[80] W. Meng, P. Lunkad, A. Kumar, K. Khayat, Influence of silica fume and polycarboxylate ether dispersant on hydration mechanisms of cement, J. Phys. Chem. C 120 (2016) 26814–26823, https://doi.org/10.1021/acs.jpcc.6b08121.