

Machine learning for high-fidelity prediction of cement hydration kinetics in blended systems

Rachel Cook^{a,*}, Taihao Han^a, Alaina Childers^a, Cambria Ryckman^a, Kamal Khayat^b, Hongyan Ma^b, Jie Huang^c, Aditya Kumar^a

^a Department of Materials Science and Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA

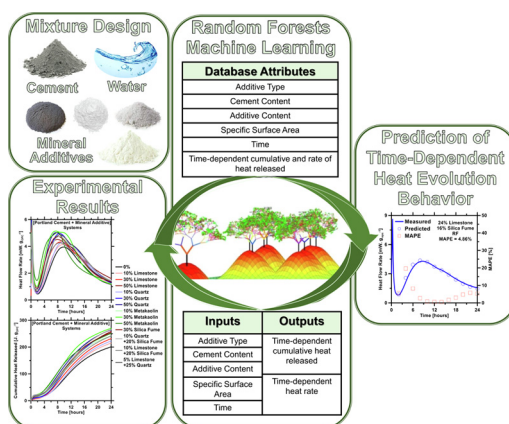
^b Department of Civil, Architectural, and Environmental Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA

^c Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA

HIGHLIGHTS

- This work contains successful prediction and optimization of Portland cement systems.
- Novel predictions of heat-evolution profiles were achieved via machine learning (ML).
- This work offers an original dataset, which contains results for 300+ unique entries.
- The database considers mixture design and physiochemical features as attributes.
- This work can be expanded to formulate mixture design based on user kinetic-criteria.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 24 October 2020

Revised 16 April 2021

Accepted 18 June 2021

Available online 21 June 2021

Keywords:

Machine learning
Random forests
Portland cement
Hydration
Mineral additives

ABSTRACT

The production of ordinary Portland cement (OPC), the most broadly utilized man-made material, has been scrutinized due to its contributions to global anthropogenic CO₂ emissions. Thus – to mitigate CO₂ emissions – mineral additives have been promulgated as partial replacements for OPC. However, additives – depending on their physiochemical characteristics – can exert varying effects on OPC's hydration kinetics. Therefore – in regards to more complex systems – it is infeasible for semi-empirical kinetic models to reveal the underlying nonlinear composition-property (i.e., reactivity) relationships. In the past decade or so, machine learning (ML) has arisen as a promising, holistic approach to predict the properties of heterogeneous materials, even without an across-the-board comprehension of the underlying composition-properties correlations. This paper describes the use of a Random Forests (RF) model to enable high-fidelity predictions of time-dependent hydration kinetics of OPC-based systems – more specifically [OPC + mineral additive(s)] systems – using the system's physiochemical attributes as inputs. Results show that the RF model can also be used to formulate mixture designs that satisfy user-imposed kinetics-related criteria. Lastly, the presented results can be expanded to formulate mixture designs that satisfy target kinetic criteria, even without knowledge of the underlying kinetic mechanisms.

© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

* Corresponding author at: Department of Materials Science and Engineering, Missouri University of Science and Technology, 205 Engineering Research Laboratory, 500 W 16th St, Rolla, MO 65409, USA.

E-mail addresses: recwx7@mst.edu (R. Cook), thy3b@mst.edu (T. Han), acff2@mst.edu (A. Childers), cr2r5@mst.edu (C. Ryckman), kkhayat@mst.edu (K. Khayat), mahon@mst.edu (H. Ma), jieh@mst.edu (J. Huang), kumarad@mst.edu (A. Kumar).

1. Introduction

Compared to other construction materials (e.g., wood; steel; masonry; etc.), concrete has the highest annual production [1], with 4.1 billion tons produced in the United States in 2019 [2]. Concrete is especially vital to emerging nations, as it is an economically viable material and its precursors plentiful enough to satisfy the demand for decent low-cost housing and infrastructure [1]. Thus, the complete eradication of concrete and cementitious materials in pursuit of a greener material is most likely an impossibility in the near future. With that stated, research that can provide improved understanding — or even further, predict the behavior and composition-property linkages — of OPC-based systems (i.e., pastes; mortars; and concretes), can improve the efficiency of the implementation of the mentioned materials in practice. As a standalone product, the production of OPC is responsible for 9% of CO₂ emissions globally [3–5]. As the overall demand for OPC continues to increase [6–8], there is rising pressure to discover alternate practices and resources to reduce CO₂ emissions resulting from OPC production [3,4,7,9]. Currently being extensively explored and optimized by researchers is partial replacement of OPC with CO₂-efficient mineral additives such as pozzolanic and filler materials. Examples of the discussed additives, in the form of ground powder, are limestone (crystalline CaCO₃), quartz (crystalline SiO₂), and silica fume (amorphous SiO₂); other less commonly used additives include polymorphs of TiO₂ (i.e., rutile and anatase), metakaolin (dominantly amorphous Al₂Si₂O₇), and corundum (crystalline Al₂O₃) [10–17]. When present in OPC-based systems, filler materials are known to alter hydration rates [10,11,18–20], typically by accelerating the hydration (i.e., the reaction with water) of the host phase by providing additional surface sites for calcium silicate hydrate (C-S-H) — universally considered to be the “glue” of OPC-based systems and the most important hydrate phase — to heterogeneously nucleate and grow upon. The acceleration of hydration rates as a result of inclusion of fillers in OPC-based systems is commonly referred to as the *filler effect* [19–21]. To follow that point, when present in OPC-based systems, pozzolanic materials are known to contribute to a pozzolanic reaction, which ultimately yields an increased percentage of C-S-H present in the system — an effect known to increase with time [18,22]. However, materials such as silica fume and metakaolin have an additional layer of complexity. These materials have been established to contribute to the hydration of OPC-systems as pozzolanic and filler materials simultaneously, with the contributions to each effect varying with time [18,22].

The hydration of the abovementioned systems is often fitted to physical results to predict the hydration behavior and microstructural development of cement systems with respect to time. A technique that is often at the forefront of cement hydration studies is isothermal calorimetry, which measures the heat absorbed or emitted, that is, *endo-/exo-thermic* processes, for a given system. The measured heat is often represented in the literature in terms of heat flow rate and the overall heat produced at a fixed temperature with respect to time: the heat flow rate is indicative of reaction rates, while the cumulative heat released can be used to extract thermodynamic information (e.g., degree of hydration of cement) at a desired time in the hydration process. The heat flow emitted over the course of the entire hydration reaction, with respect to each unique cementitious system, can serve as a characteristic heat-evolution “fingerprint,” yielding information regarding underlying kinetic mechanisms. Early hydration of cement is often described in four periods that correspond to observable regime changes in heat flow rate as a function of time. The four stages of early hydration are often referred to in the literature [23] as: the (I) *initial period*; (II) *induction period*; (III) *acceleration period*; and (IV) *deceleration period*.

As stated above, these kinetic phenomena are commonly observed via unique heat evolution signatures with time, which are used to demonstrate hydration degree and microstructural development of cement systems, typically by numerical kinetic models [24–27]. Arguably the most prominent, prolifically-cited numerical kinetic works developed in the past century, relatively around the same time, has been by William A. Johnson and Robert F. Mehl [28], Melvin Avrami [29–31], and Andrey N. Kolmogorov [32], whose collective work is commonly denoted as JMAK kinetics. JMAK kinetics assume that nucleation occurs in a random, homogeneous fashion over the total untransformed material of the given system, where the growth rate is an independent factor and assumed to occur isotropically. Several studies [33–40] have applied JMAK kinetics to cement hydration, with relatively poor fits. For some time, it's been known that such assumptions do not fit with what has been experimentally observed in terms of cement microstructure development after mixing. That is, it is well established that one of the more important hydration products of OPC systems, C-S-H, grows heterogeneously in a needle-like fashion (sometimes also referred to as fibrillar [41], sheet-like [42], or globular [43] in the literature) on cement surfaces at early ages, eventually leading to setting and the consequent development of mechanical properties. Thomas [44] pointed out that experimental observations contradict assumptions set by the JMAK equations and consequently chose to frame cement hydration in the context of John W. Cahn's original work [45]. Cahn's boundary nucleation and growth model — similar to the JMAK equations — is based on a few assumptions; Cahn's work assumes that nuclei form on planar boundaries that are randomly oriented and distributed within the system, a constant nucleation rate per unit area of the unreacted surface, and a constant, isotropic growth rate.

Cahn's work has served as a springboard for several numerical studies [16,18,19,21,22,24,44,46–57], whose collective work has sought to comprehensively explain the rate-limiting steps driving the early stages of hydration — corresponding to nucleation and growth — and eventually leading to the slowing or decelerating of hydration reaction rates. Though the referenced works [16,18,19,21,22,24,44,46–57] serve as examples of important milestones in understanding underlying hydration mechanisms of OPC-based systems, there are still points of contention within the literature. For example, in regards to the aforementioned slowing of hydration rates that occurs 10 hours or so after mixing for a plain OPC-based system, there is debate on whether the *deceleration period* is rate-limited by diffusion-limited [16,50–53] or dissolution-limited [54,58,59] kinetics. In addition, the process of elucidating hydration mechanisms to then predict hydration behavior based on the aforementioned mechanisms is an approach that has and still requires numerous studies with varying physical and numerical approaches.

Currently within the literature, there are numerical methods that can predict properties based on physical data, however these methods do not consider reaction kinetics, but instead take an engineering approach by utilizing artificial intelligence, machine learning (ML) techniques. Previous studies [60–69] have proposed that improving understanding of the relationships between hydration kinetics and consequent mechanical properties [60–66,70–79] of OPC-based systems can be assisted by a ML approach, based on the large-scale analysis of experimental data. Bangaru et al. [69] applied Random Forests (RF), Naïve Bayes, Logistic Regression, K-Nearest Neighbors, and Support Vector Machine models to predict the degree of hydration based on the microstructural development of concrete systems. Conversely, Cruz et al. [68] predicted the microstructural development of cement systems using the degree of hydration as an input via an Artificial Neural Network model. These studies [68,69] reveal that ML can serve as a promising plat-

form to predict the heat evolution of the hydration reaction, which can be directly linked to the kinetics of the overall reaction. By utilizing ML platforms, complexities pertaining to large compositional degrees of freedom (i.e., mixture design variables, permutations of which can vary significantly and exert substantial influence on properties) and the consequent, nonlinear relationships between design variables and properties of OPC-based systems can be overcome. For example, the inclusion of ground limestone in OPC-based systems, depending on the replacement level, can lead to formation of additional, carboaluminate phases in OPC-based systems via reactions with alumina-containing anhydrous phases (e.g., C_3A) [14,80–83], or from destabilization of the monosulfoaluminate phase [84]. The dissolution of metakaolin is known to release aluminate $[Al(OH)_4^-]$ anions, ultimately suppressing the nucleation and growth of C–S–H [18]. Further, mineral additives, as stated previously, can function as pozzolanic or filler materials or both to varying degrees at different ages, complicating the overall hydration process. Their performances have been linked to the available specific surface area (SSA; $cm^2.g^{-1}$) and other parameters related to physiochemical effects, such as agglomeration [18,22], which can effectively alter the total SSA contributing to the hydration reaction. Therefore, sophisticated approaches, such as ML, are required to reveal the hidden, and complex, semi-empirical rules that govern the correlation between mixture design and properties of OPC-based systems.

In this study, the RF model — a modification of the classification-and-regression decision trees (CART) ML models — is used to perform novel predictions of the time-dependent, kinetically-related heat-evolution behavior with variations according to different mineral additive types, such as quartz, limestone, metakaolin, and silica fume and physiochemical attributes such as SSA. The prediction results show that the RF models can predict and optimize the relatively continuous (i.e., short time steps) and long time period (i.e., 24 h) heat-evolution-determined kinetic profiles corresponding to plain and [OPC + mineral additive] systems as well as predict profiles for new systems when properly and rigorously trained, a feat that is currently impossible with current numerical kinetic models. The database constructed from heat evolution experimental data includes 1-additive and 2-additive systems. In order to evaluate the performance of the ML model, five different statistical parameters are used to compare the predicted cumulative heat and heat flow rate against measured data. The correlations between the inputs and outputs developed by the RF algorithms utilized in this study can be used to optimize the mixture design based on desired hydration kinetics. Further, by utilizing the reactivity of cementitious systems in both the differential (i.e., heat flow rate) and integral (i.e., cumulative heat released) forms, this study shows that ML can be applied regardless of measurement technique and data form (i.e., differential or integral).

2. Overview of the Random Forests (RF) model

The RF model — a modification of classification-and-regression-trees (CART) model, as detailed in our previous studies [60,61,85–89] — constructs a large number of uncorrelated CART trees as a committee to produce independent outputs and ultimately averages them to produce the final output [90]. Each tree within the RF algorithm is constructed via binary splits into “near-homogeneous” terminal nodes; such splits are done in a recursive fashion until the optimal structure of the tree is achieved. The RF algorithm leverages the technique of bagging [91,92], which ensures that each tree grows from a randomly-selected group of bootstrap samples, each comprising of the same number of inputs as the entire training dataset. The RF algorithm also leverages another technique, bootstrapping, which helps to reduce the variation (underfitting) and bias (overfitting) among the 100 s-to-1000 s of trees that are grown in the forest [93]. Another advantage of the RF model — compared to other ML

platforms — is that it allows each tree to grow to its maximum size without any smoothing or pruning whatsoever; this helps maintain diversity among the trees (i.e., output of each tree is truly independent of the output of all other trees), thereby allowing the model to not just capture trends in the dataset, but also account for outliers. Lastly, the RF model contains two hyper-parameters that need to be adjusted manually to achieve the optimal prediction. These hyper-parameters — that is, the *number of trees in the forest* and *number of splits in each tree* — were optimized in this study using the 10-fold cross validation (CV) method [61,62,85,87–89,94,95] in conjunction with a grid-search method [96] that is described in Section 4.2.

3. Database collection and assessment of prediction accuracy of ML model

The cumulative heat and heat flow rate corresponding to the hydration of [OPC + mineral additive] paste systems collected from isothermal calorimetry, described in the *Supplementary Information* document, were consolidated into the training database (Table 1) and the testing database (Table 2). The first-mentioned database was used for training the ML model, and subsequently the testing database was used for evaluating its prediction performance (i.e., ability to predict hydration heat flow rate and cumulative heat that were precluded from the training database). The training database is comprised of 7800 unique data-records from 326 systems, wherein the time-dependent cumulative heat and heat flow rate of each mixture design at every hour from 0 to 24 hours are outputs. The training database includes 8 inputs related to physicochemical properties of the system: mineral additive type (e.g., 0 = OPC; 1 = Quartz; 2 = Limestone; 3 = Metakaolin; 4 = Silica Fume; 5 = Quartz + Limestone; 6 = Quartz + Metakaolin; 7 = Quartz + Silica Fume; 8 = Limestone + Metakaolin; 9 = Limestone + Silica Fume; 10 = Metakaolin + Silica Fume); mineral additive type (Unitless); normalized OPC content (Unitless); normalized additive-1 content (Unitless); normalized additive-2 content (Unitless); SSA of OPC ($cm^2.g^{-1}$); SSA of additive-1 ($cm^2.g^{-1}$); SSA of additive-2 ($cm^2.g^{-1}$); and time (hour). For both 1-additive and 2-additive systems, the mineral additive replacement level varied as described in the *Supplementary Information* document. Statistical parameters pertaining to the training database are shown in Table 1. The testing database consisted of 312 unique data-records from 26 systems and the same 8 inputs and 2 outputs as the training database. The time-dependent cumulative heat and heat flow rate of each system were predicted every two hours from 0 to 24 hours. The mineral additive replacement levels for both 1-additive and 2-additive system in the testing database were randomly selected. Statistical parameters pertaining to the testing database are shown in Table 2.

In this study, 5 unique statistical parameters — Pearson correlation coefficient (R); mean absolute percentage error (MAPE); coefficient of determination (R^2); root mean squared error (RMSE); and mean absolute error (MAE) — were to quantitatively and rigorously assess the prediction performance of a ML model (RF) against the testing databases. Mathematical formulations for each of these parameters can be found elsewhere [60,85].

4. Results and discussion

4.1. Experimental isothermal calorimetry of [OPC + mineral additive] systems

Isothermal calorimetry was used to measure the hydration rates of OPC that was partially replaced by four mineral additive types (e.g. quartz, limestone, metakaolin, and silica fume), in

Table 1

Summary of four statistical parameters related to each of the 10 attributes (8 inputs and 2 outputs) of the training database. The training database included 8 inputs related to physicochemical properties of the system: mineral additive type (e.g., 0 = OPC; 1 = Quartz; 2 = Limestone; 3 = Metakaolin; 4 = Silica Fume; 5 = Quartz + Limestone; 6 = Quartz + Metakaolin; 7 = Quartz + Silica Fume; 8 = Limestone + Metakaolin; 9 = Limestone + Silica Fume; 10 = Metakaolin + Silica Fume); mineral additive type (Unitless); normalized OPC content (Unitless); normalized additive-1 content (Unitless); normalized additive-2 content (Unitless); SSA of OPC ($\text{cm}^2\cdot\text{g}^{-1}$); SSA of additive-1 ($\text{cm}^2\cdot\text{g}^{-1}$); SSA of additive-2 ($\text{cm}^2\cdot\text{g}^{-1}$); and time (hour). The training database additional includes 2 outputs, that is, the time-dependent cumulative heat and heat flow rate of each system were predicted every two hours from 0 to 24 hours. The database consists of 7800 unique data-records.

Attribute	Unit	Min.	Max.	Mean	Std. Dev.
Mineral Additive Type	Unitless	0.0000	10.000		
Normalized OPC Content	Unitless	0.2650	0.6902	0.4670	0.1333
Normalized Additive-1 Content	Unitless	0.0000	0.4136	0.1694	0.1105
Normalized Additive-2 Content	Unitless	0.0000	0.3787	0.0507	0.0856
Cement SSA	$\text{cm}^2\cdot\text{g}^{-1}$	1726.5	1726.5	1726.5	0.0000
Additive-1 SSA	$\text{cm}^2\cdot\text{g}^{-1}$	0.0000	198,000	22,659	43,392
Additive-2 SSA	$\text{cm}^2\cdot\text{g}^{-1}$	0.0000	198,000	34,359	71,681
Time	hour	1.0000	24.000		
Cumulative Heat	$\text{J}\cdot\text{g}_{\text{opc}}^{-1}$	4.1620	312.53	141.75	83.571
Heat Flow	$\text{mW}\cdot\text{g}_{\text{opc}}^{-1}$	0.4117	6.0298	2.7629	1.2437

Table 2

Summary of four statistical parameters related to each of the 10 attributes (8 inputs and 2 outputs) of the testing database. The database consists of 312 unique data-records.

Attribute	Unit	Min.	Max.	Mean	Std. Dev.
Additive Type	Unitless	0.0000	10.000		
Normalized Cement Content	Unitless	0.2827	0.6132	0.4169	0.1186
Normalized Additive-1 Content	Unitless	0.0690	0.3714	0.1961	0.0908
Normalized Additive-2 Content	Unitless	0.0000	0.3171	0.0758	0.0955
Cement SSA	$\text{cm}^2\cdot\text{g}^{-1}$	1726.5	1726.5	1726.5	0.0000
Additive-1 SSA	$\text{cm}^2\cdot\text{g}^{-1}$	1054.1	198,000	32,552	60,228
Additive-2 SSA	$\text{cm}^2\cdot\text{g}^{-1}$	0.0000	198,000	49,603	81,587
Time	hour	2.0000	24.000		
Cumulative Heat	$\text{J}\cdot\text{g}_{\text{opc}}^{-1}$	16.407	301.20	161.21	84.608
Heat Flow	$\text{mW}\cdot\text{g}_{\text{opc}}^{-1}$	0.5321	7.2012	2.8690	1.2483

1-additive (Fig. 1) and 2-additive designs (Fig. 2). Figs. 1 and 2 show representative heat evolution profiles of OPC paste systems prepared by substituting OPC with different mineral additives at various replacement levels. In Fig. 1, all additive replacements appear to substantially enhance OPC hydration rates, by means of a leftward shift of the main hydration peak – indicated by a shortened induction period – and higher heat flow rate peak. The acceleration of hydration rates increases with increasing additive replacement level (Fig. 1a) due to the filler effect [19–21] – a phenomenon which intensifies as the total solid surface area of the system increases.

Fig. 1b and 1c compare the heat evolution profiles of OPC replaced by the four mineral additives, with 1-additive included in each system, at the identical replacement level of 20 wt%. Each system exhibits accelerated hydration rates compared to the neat system, largely due to the filler effect [18–22]. However, the varying degrees of acceleration (Fig. 1b) and cumulative heat released at 24 hours (Fig. 1c) are a result of the varying chemical and physical effects, which are consequent of the inclusion of each mineral additive, as discussed in Section 1. These systems (shown in Fig. 1) are fairly simple and have been predicted using traditional thermo-kinetic models [18,19,21,22,57].

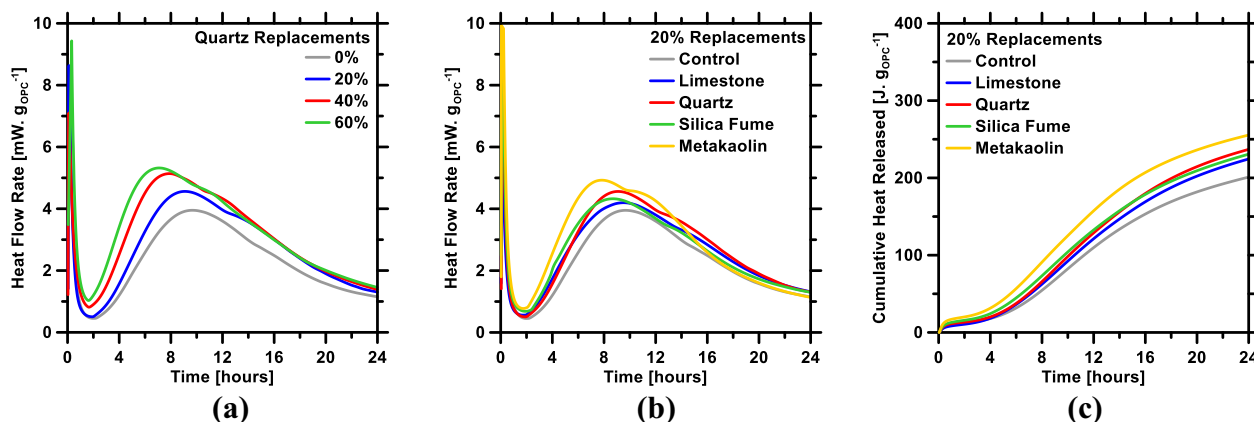


Fig. 1. Isothermal calorimetry determinations of time-dependent (a) heat flow rate of [OPC + quartz] systems at replacement levels varying from 0 wt% to 60 wt%, and (b) heat flow rate and (c) cumulative heat release of [OPC + 1-additive] systems at identical 20 wt% replacements of limestone; quartz; silica fume; and metakaolin. All calorimetry measurements were recorded over the initial 24 hours of hydration. Uncertainty in heat flow rate at the main hydration peak is $\pm 2\%$.

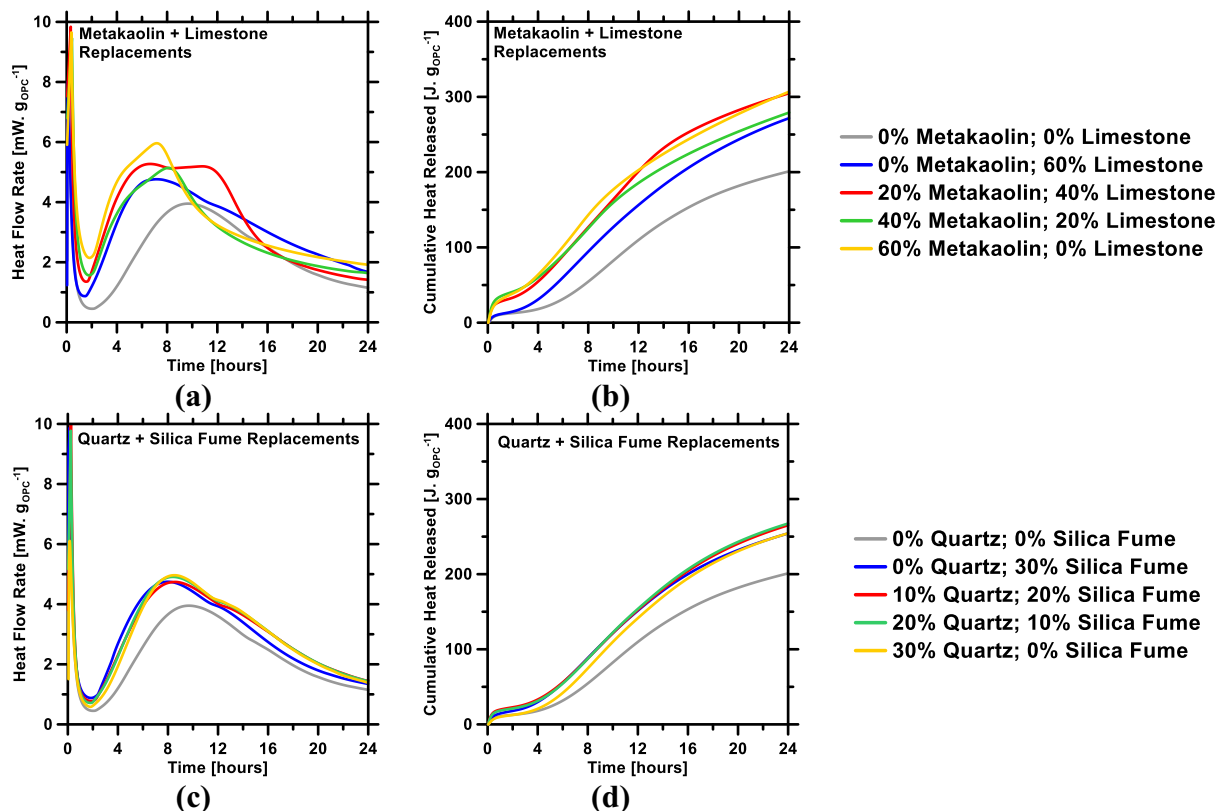


Fig. 2. Isothermal calorimetry determinations of time-dependent 2-additive systems, that is, (a) heat flow rate and (b) cumulative heat released of [OPC + metakaolin + limestone] systems as replacement levels ranging from 0% to 60% with respect to each additive in the system; and (c) heat flow rate and (d) cumulative heat released of [OPC + quartz + silica fume] systems as replacement levels ranging from 0% to 30% with respect to each additive in the system. All calorimetry measurements were recorded over the initial 24 hours of hydration. Uncertainty in heat flow rate at the main hydration peak is $\pm 2\%$.

However, for only slightly more-complex OPC systems (i.e., OPC systems prepared by replacing OPC with 2-additives at different replacement levels), such as those represented by the heat evolution profiles shown in Fig. 2, the hydration process of each system is much more difficult to predict by traditional means due to increased compositional degrees of freedom [68,69]. [OPC + metakaolin + limestone] systems (Fig. 2a and 2b), although showing accelerated heat flow rates and greater cumulative heat released at 24 hours compared to the neat system, display no clear trend (s) in terms of the occurrence of main hydration peak and the cumulative heat released at 24 hours with increasing replacement levels. With the many behaviors metakaolin is known to exhibit in a simple system [18,57], it is difficult to determine which mechanism dominates its contributions to the acceleration of OPC hydration. Adding another layer of complexity is the incorporation of limestone, which can form carboaluminate phases in the presence of aluminates phases [14,80–84]. Heat evolution profiles corresponding to [OPC + quartz + silica fume] systems, (Fig. 2c and 2d) also exhibit no clear trends with increasing replacement level, with the exception of accelerated heat flow rates and greater cumulative heat released at 24 hours compared to the neat system. However, it can be speculated that the *filler effect* is the dominating effect in [OPC + quartz + silica fume] systems based on the observations from Fig. 2 as well as these related works [10,11,18–22]. Even so, it is truly unknown if the accelerated rates resulting from the incorporation of quartz and silica fume in this study are equivalent or not. The contributions of additive chemistry and/or physical effects such as effective SSA can potentially vary with both the individual replacement level and total replacement level. Such complexity, nonetheless, is expected because each input variable

— pertaining to either the additive chemistry, mixture design, or SSA — consistently casts unique and significant impact on the OPC hydration; when more than one input variable are concomitantly adjusted, especially hydration mechanism(s), the cumulative impact on properties is even more complex. Precisely because of such complexities, derivation of empirical, concurrent physical and chemical property relationships in OPC hydration is not feasible using simple statistical and analytical kinetic models; more sophisticated algorithms such as machine learning are needed for such tasks.

4.2. Heat evolution prediction of [OPC + mineral additive] systems

As described in Section 3, the RF model was trained using the training database; thereafter, the prediction performance of the trained ML model was evaluated against the testing database. To maximize the RF model's prediction performance, it is important to ascertain that: inputs-output correlations are properly established; outliers are accounted for; and variance and bias among trees (i.e., CARTs) of the model are kept as low as possible. To accomplish these objectives, the two hyper-parameters of the RF model (i.e., number of trees in the forest; and number of splits per tree) were rigorously optimized based on the nature and volume of the database. In this study, for such optimizations, the grid-search method [96,97] was used. This method involves autonomous, iterative variations in the hyper-parameters — while concurrently employing the 10-fold CV method [94] — to determine optimal values of hyper-parameters that result in a minimum deviation between RF predictions and measured values. The aforesaid deviation between predictions and observations is quantified using

all five statistical parameters listed in Section 3 (i.e., R ; R^2 ; MAE; MAPE; and RMSE). Simply put, hyper-parameters are determined to be optimal when R and R^2 are at (or close to) their global maximum, while MAE, MAPE, and RMSE are at (or close to) their global minimum.

The representative results of the RF model, which were obtained from the grid-search method, are shown in Fig. 3. Two of the statistical parameters – that is, MAPE and R^2 – that were used to measure the deviation between predictions and measured values of cumulative heat flow and heat flow rate (averaged over the initial 24-hour period) are showcased. On the basis of MAPE and R^2 , the optimal prediction performance of the RF model for the cumulative heat prediction occurred for common values of the two hyper-parameters: that is, *number of trees in the forest* = 800; and *number of splits in each tree* = 5. Moreover, the RF model was structured using 800 trees and 5 splits produced accurate prediction of the heat flow rate as well. Therefore, 800 trees and 5 splits were selected as optimal hyper-parameters for further predictions in this study. When the *number of splits* was less than 5, logical splits in the databases were numerically inadequate and too simplistic to fully encompass the complex, underlying correlations between inputs and output. When the *number of splits* was larger than 5, the complex structure of the trees (CARTs) heightened the likelihood of bias, which in turn resulted in overfitting. Likewise, when the *number of trees* was less than 800, the RF model did not have enough independent bootstraps to produce accurate predictions (for new OPC systems in the testing dataset). However, when excessive trees (i.e., greater than 800) were used, the predic-

tion performance did not improve (akin to law of diminishing returns [91,98]), even though the computational complexity of the model indubitably increased. This is hypothesized to be caused by increased redundancy among the trees. More specifically, it is expected that – in a forest with greater than 800 trees, all of which were meant to be distinct – several trees (that were forced to be grown from similar bootstraps) ended up having similar structures, and, therefore, produced similar predictions; thereby, resulting in little to no improvement in the RF model's overall prediction accuracy.

Prediction performance using heat evolution data, related to the kinetics of the hydration reaction corresponding to [OPC + mineral additive] systems, are compared – in the form of the five statistical parameters – against extracted physical values in Figs. 4–6 for every two hours. The statistical parameters (averaged over the initial 24-hour period of hydration) corresponding to the testing set are itemized in Table 3. Fig. 6 shows representative predicted results against the measured values; for reference, the entire cumulative heat and heat flow rate spectrum of representative systems from calorimetry experiments are included to exhibit a visual comparison between predicted values and measured values.

As can be seen in Table 3, predictions of heat flow rate and cumulative heat, as produced by the RF model, are accurate. The prediction of heat flow rate had a Pearson correlation coefficient (R) value of 0.965 and a root mean squared error (RMSE) value of 0.331 $\text{mW} \cdot \text{g}_{\text{opc}}^{-1}$, while the prediction of the cumulative heat had a Pearson correlation coefficient (R) value of 0.989 and a root mean squared error (RMSE) value of 14.398 $\text{J} \cdot \text{g}_{\text{opc}}^{-1}$. In Figs. 4 and 5, it is

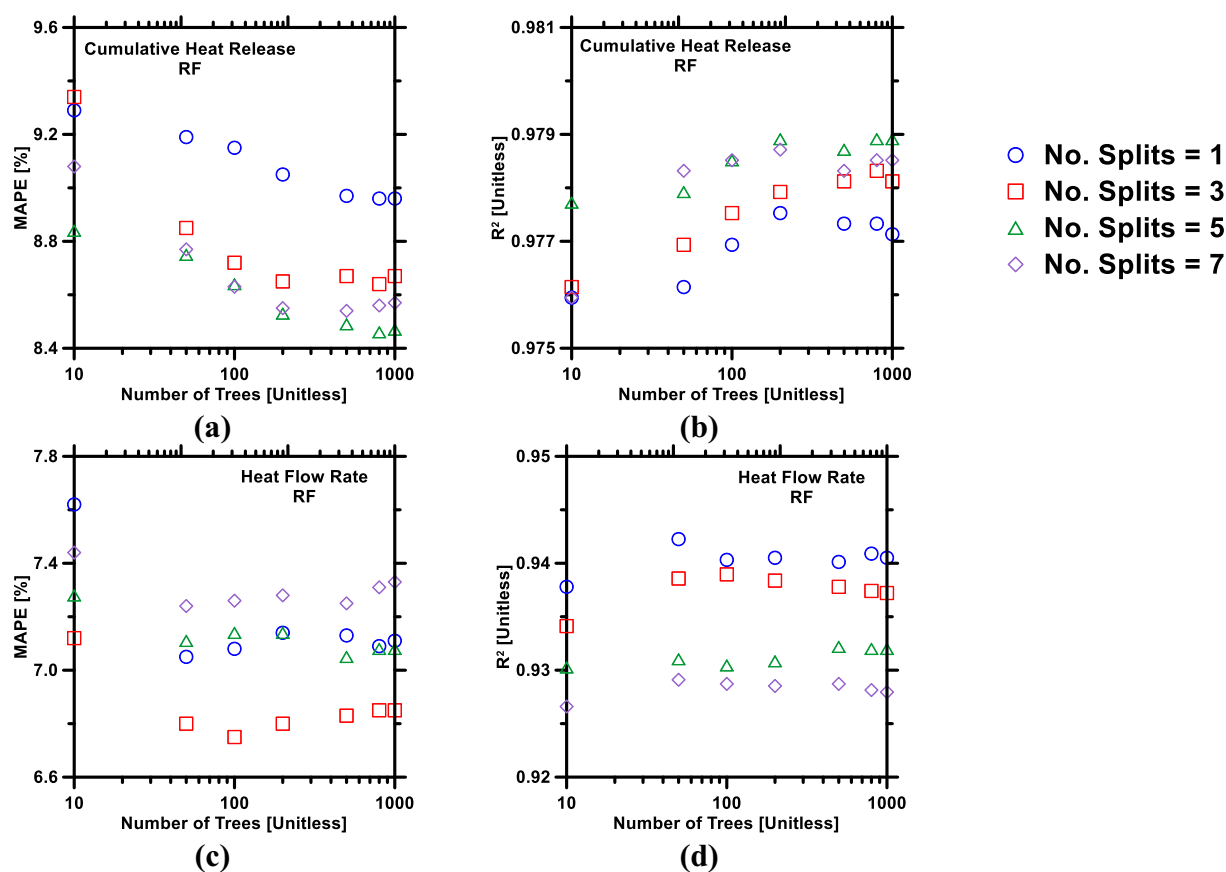


Fig. 3. Grid-search method used to optimize hyper-parameters (*number of trees in the forest*; and *number of splits in each tree*) of the RF model to improve its prediction performance against: (a) cumulative heat flow, as evaluated by MAPE; (b) cumulative heat flow, as evaluated by R^2 ; (c) heat flow rate, as evaluated by MAPE; and (d) heat flow rate, as evaluated by R^2 .

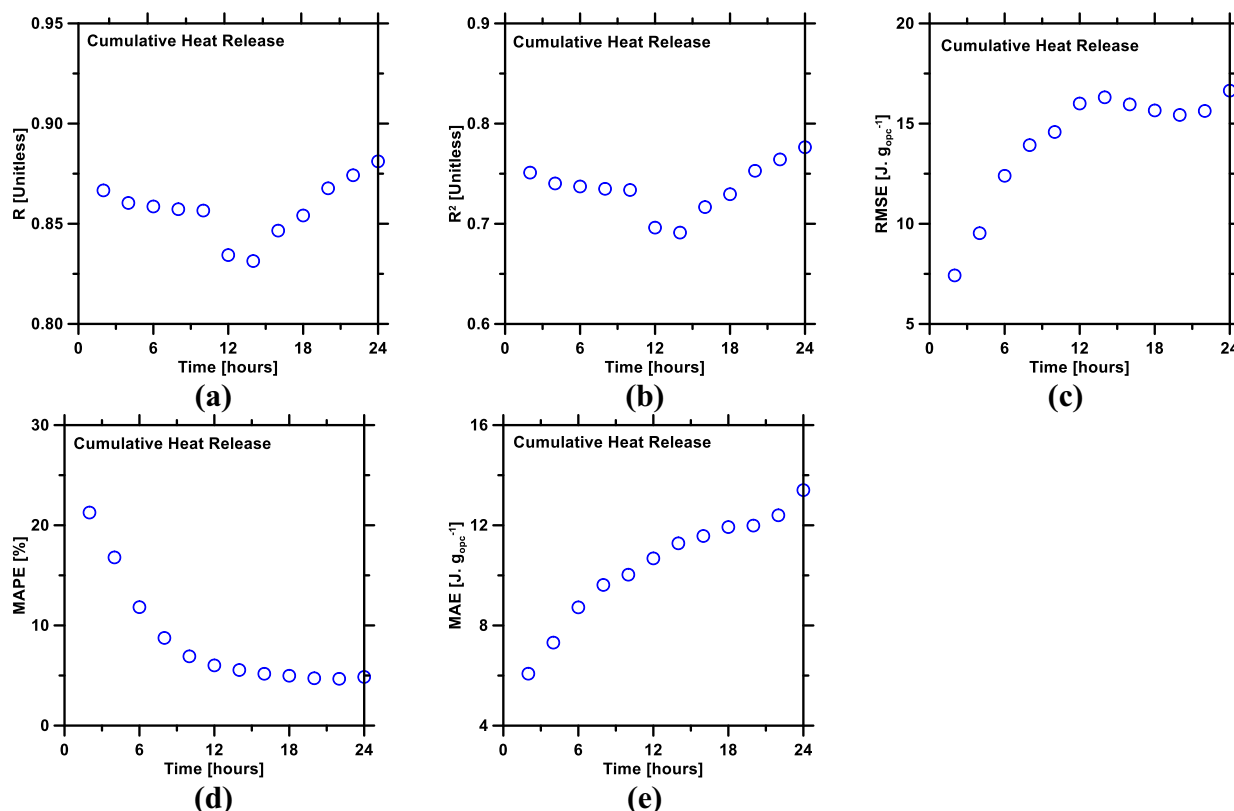


Fig. 4. Statistical parameters describing errors in prediction of cumulative heat release of pastes over a 24-hour period in the testing database: (a) Pearson correlation coefficient (R); (b) coefficient of determination (R^2); (c) root mean squared error (RMSE); (d) mean absolute percentage error (MAPE); and (e) mean absolute error (MAE) as functions of time.

worth pointing out that the prediction for early-age hydration reaction behavior reveals lower accuracy than the prediction for later hours because of significant variations among pastes. For example, the hydration reaction needs to experience three stages – the *initial period*, *induction period*, and *acceleration period* – within the first 2-to-4 hours, and each stage exhibits a unique “footprint” (e.g., distinct kinetic behaviors) corresponding to different mechanisms. However, hydration undergoes only two stages (i.e., the *acceleration period* and *deceleration period*) for the latter twenty hours, during which the kinetic behavior is occurring at a significantly slower rate. As seen in Fig. 6 (g–h), the prediction of the paste containing metakaolin replacement presented relatively lower accuracy compared to other [OPC + mineral additive] systems without metakaolin. The most plausible reason is that the volume of the training database used is not large enough, despite containing more than 300 unique compositions (i.e., OPC systems). The inclusion of more data-records into the database will enhance its volume and diversity, which will, in turn, further reduce the model's prediction errors, especially errors for the metakaolin system, thereby making the model more amenable for optimization-based tasks.

The high-fidelity prediction of cumulative heat and heat flow rate from the RF model is expected because several past studies have already reported that the RF model produces superior predictions of materials' properties [60–62,78,85,86,99]. This disparity in the RF model's prediction performance vis-à-vis other ML model can be traced back to the former model's structure, which gives it several advantages [91,92,100]. In the RF model, a large number of trees (i.e., number of CARTs $\gg 100$) are grown, one-by-one in a recursive manner by using randomly-selected bootstraps of identical volume; as such, generalization errors (likelihood of overfitting)

are minimized [93]. As each tree is permitted to grow – and not pruned or smoothened at all – until it reaches its maximum size, the RF model is proficient at developing rational *input–output* correlations, while ensuring that seemingly anomalous data-records (i.e., outliers with respect to already established trends) are not ignored or removed during any stage of the training process. Furthermore, the RF model employs two-stage randomization, which goes a long way in ensuring that each of the deep, unpruned trees, is distinct in its structure and does not exhibit any dependency to the rest of the trees in the forest. Such independency among the trees is crucial because it ensures that predictions produced by the trees are truly independent of each other; which in turn, results in low variance in the final predictions (i.e., average of predictions from all trees). Lastly, the RF model is easy to implement because the *number of trees in the forest* and *number of splits in each tree* are the only two hyper-parameters that are required as inputs from the user. Adjusting these parameters through trial-and-error is generally cumbersome and time-consuming, and could compromise prediction performance [94]. Therefore, in this study, we used the grid-search method along with the 10-fold CV method (see Fig. 3) for such adjustments.

4.3. Optimization of [OPC + mineral additive] mixture design

The results and discussion shown in the above section have proven that the RF model, and further ML in general, can be utilized to predict the time-dependent heat flow rate and cumulative heat corresponding to the hydration reaction of [OPC + mineral additive] systems – in relation to additive replacement level and particle size distribution – in a high-fidelity manner. The authors, therefore, posit that this ability of the RF model – to understand

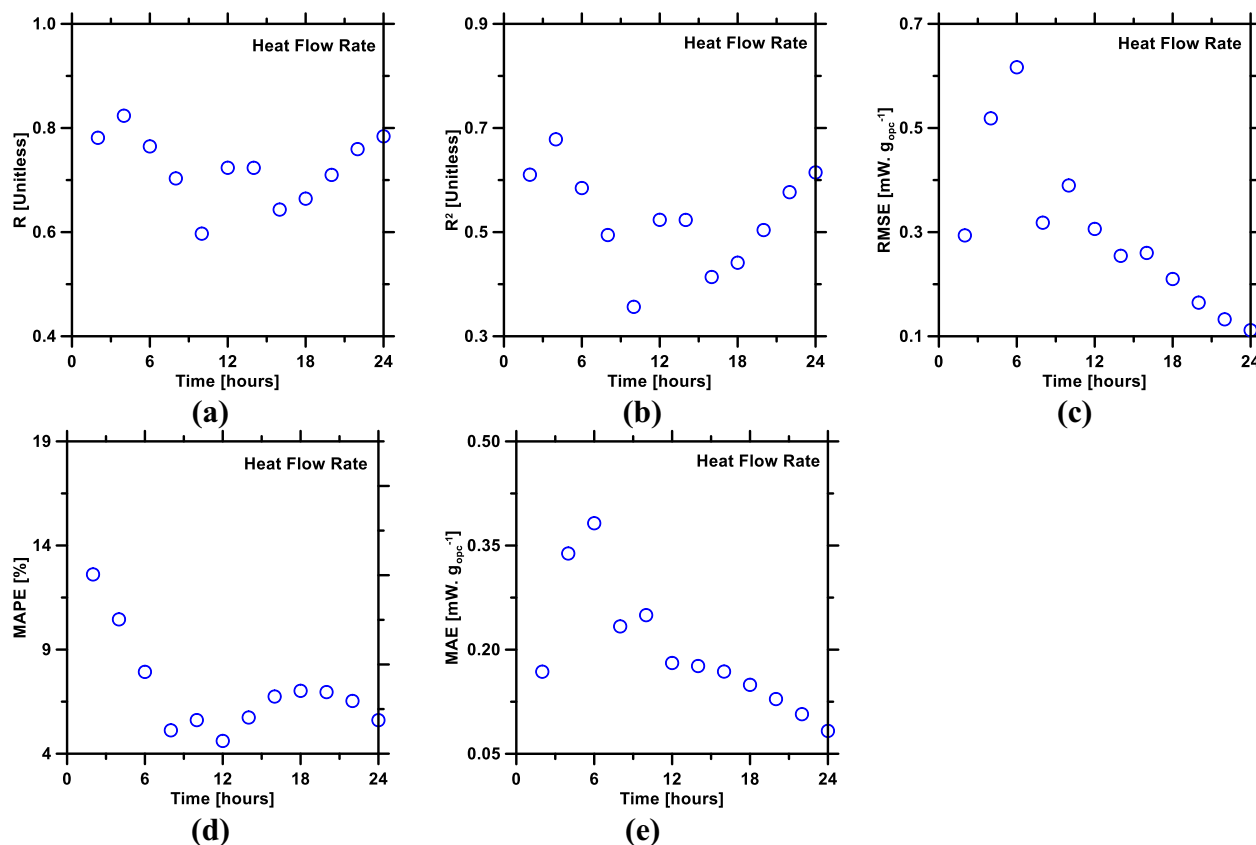


Fig. 5. Statistical parameters describing errors in prediction of heat flow rate of pastes over a 24-hour period in the testing database: (a) Pearson correlation coefficient (R); (b) coefficient of determination (R^2); (c) root mean squared error (RMSE); (d) mean absolute percentage error (MAPE); and (e) mean absolute error (MAE) as functions of time.

hidden correlations between physiochemical attributes of the paste and hydration behavior of cement in such pastes – can be leveraged to develop optimal mixture design of [OPC + mineral additive] systems that exhibit target (user-imposed) hydration behavior. To verify this, an optimization scheme was designed to accept three target calorimetric features – the heat flow rate corresponding to the main hydration peak; time corresponding to the occurrence of the main hydration peak; and the slope of heat flow curve during the acceleration period – as inputs, along with selected mixture design inputs. Then, the Bayesian optimization approach [101,102] was employed, wherein the RF model was invoked to leverage its knowledge of composition-reactivity correlations (obtained during its training and validation) to reveal the optimal mineral additive type, additive SSA, and additive replacement level, which in combination would yield a calorimetry profile featuring the target (user-imposed) calorimetric features (i.e., slope of acceleration; time of main hydration peak; and heat flow rate at the main hydration peak). To the best of authors' knowledge, no kinetic models reported in previous studies are capable of producing such reverse predictions of the mixture design (i.e., mineral additive type; SSA; and replacement level) using heat evolution signatures as inputs.

During the optimization process, four variables were used as the primary inputs: the heat flow rate corresponding to the main hydration peak ($J_{g_{opc}}^{-1}$); time corresponding to the occurrence of the main hydration peak (hour); slope of the acceleration period ($J_{g_{opc}}^{-1} \cdot \text{hour}^{-1}$); and normalized water mass (unitless). The optimization process – altogether, comprised of three steps – was utilized to determine physicochemical information corresponding to the relevant mineral additive, which were subsequently used

to achieve the targeted heat flow rate signature corresponding to 1-additive systems. (I) Initially, the RF model was used to predict the mineral additive type (unitless) present in each cement system using the aforementioned four inputs. (II) The output mineral additive type was then utilized as an additional, secondary input (i.e., in addition to the primary inputs) to consecutively predict the mineral additive's SSA ($\text{cm}^2 \cdot \text{g}^{-1}$). (III) In the final step, both the mineral additive type and the SSA of the mineral additive, predicted in steps (I) and (II), respectively, were utilized as additional, secondary inputs adjunct to the primary inputs to predict the replacement level of the relevant mineral additive. The training dataset consisted of 208 1-additive systems; while the target calorimetric parameters were extracted from six 1-additive systems (see Fig. 7) that were randomly selected from the testing dataset used in the above section. Results obtained from the optimizations, corresponding to the 1-additive systems, are shown in Table 4. Here, the optimal values of mineral additive type, additive replacement level, and additive's SSA – as produced by the RF model – are compared against actual values.

As can be seen in Table 4, the optimization of mineral additive type was successful, and the predictions of additive SSA and replacement level are reasonably accurate. Overall, these results are expected. This is because various combinations of replacement level and effective SSA of additives can yield similar heat flow behavior, thereby leading to potential error (i.e., deviation between predicted and actual additive replacement level or SSA). For example, high peak heat flow rate could be obtained by using a coarse mineral additive at high replacement level, or a fine mineral additive at low replacement level. The presence of different mineral additives in each OPC system contributes to the uniqueness of

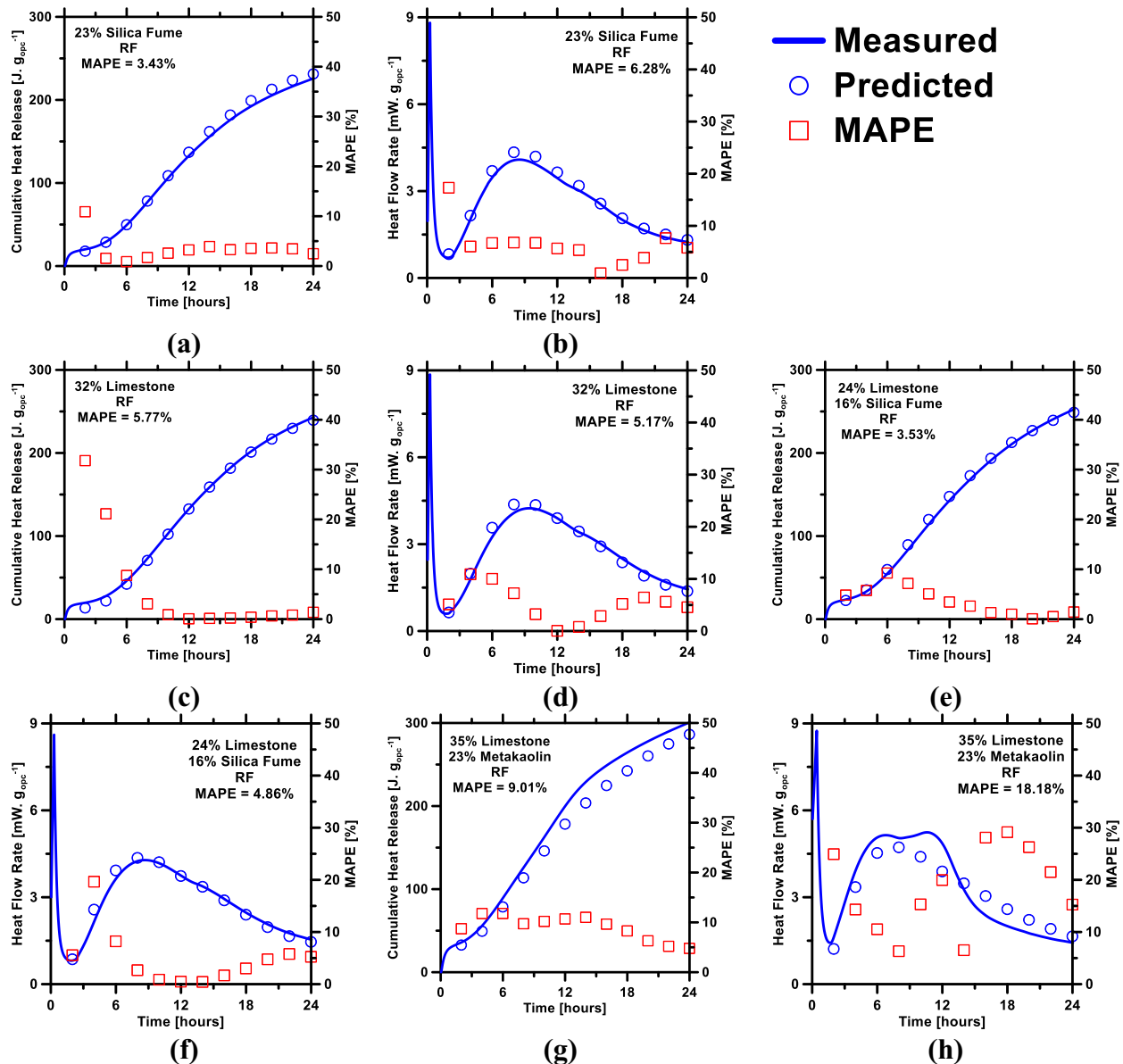


Fig. 6. The RF model's predictions of: (a) cumulative heat and (b) heat flow rate of [OPC + silica fume]; (c) cumulative heat and (d) heat flow rate of [OPC + limestone]; (e) cumulative heat and (f) heat flow rate of [OPC + silica fume + limestone]; and (g) cumulative heat and (h) heat flow rate of [OPC + metakaolin + limestone] compared against experimental measurements. Mean absolute error (MAE) of each prediction is shown in the figure. The overall mean absolute errors (MAE) of the predictions are shown in the legends.

Table 3

Prediction performance of ML models, measured on the basis of the cumulative heat release and heat flow rate of cement pastes in the testing database. Five statistical parameters (i.e., R , R^2 , MAE, MAPE, and RMSE) – averaged over the initial 24-hour period – are shown.

	ML Model	R	R^2	MAE	MAPE	RMSE
Heat Flow Rate	RF	<i>Unitless</i>	<i>Unitless</i>	$\text{mW.g}_{\text{opc}}^{-1}$	%	$\text{mW.g}_{\text{opc}}^{-1}$
		0.965	0.932	0.197	7.078	0.331
Cumulative Heat Release	RF	<i>Unitless</i>	<i>Unitless</i>	$\text{J.g}_{\text{opc}}^{-1}$	%	$\text{J.g}_{\text{opc}}^{-1}$
		0.989	0.979	10.417	8.460	14.398

the respective heat flow signatures, consequent of the hydration reaction. Thus, additive type is highly predictable when ML techniques are properly utilized.

A limitation of the optimization scheme described above is insufficient data-records (i.e., 208 unique compositions), and limited variations in the inputs, leading to error in predictions of replacement level and SSA of the additive. A larger and

more diverse database will enable the ability to optimize mixture design in high-fidelity manner. With that stated, there is potentially room for improvement within the optimization process in future studies; however, the novel work within this publication is the first step to optimize cementitious mixture designs that are likely to exhibit target/desired heat evolution signatures.

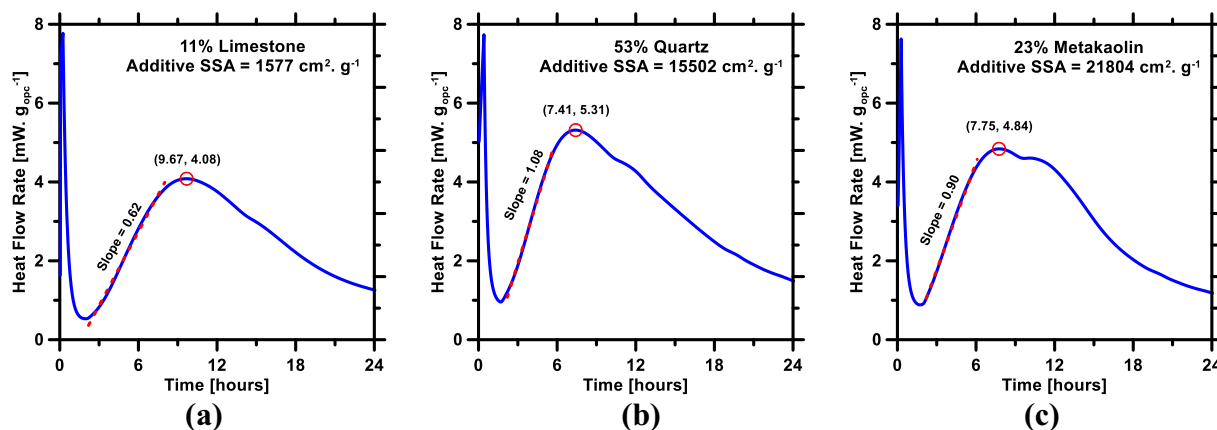


Fig. 7. Isothermal calorimetry profiles of (a) [OPC + limestone]; (b) [OPC + quartz]; and (c) [OPC + metakaolin]. These, and heat evolution profiles of three other 1-additive systems, were randomly selected from the testing database. From these profiles, heat flow rate and time of the main hydration peak (marked by red circle), and slope of the acceleration period (marked by red dashed line) were extracted; and subsequently used as inputs (target calorimetric parameters) for mixture design optimization.

Table 4

Optimization results of six 1-additive systems compared against actual values of additive type, additive replacement level, and additive SSA.

	11% Limestone			53% Quartz			23% Metakaolin		
	Actual	Predicted	Error (%)	Actual	Predicted	Error (%)	Actual	Predicted	Error (%)
Additive Type	2	2	0	1	1	0	3	3	0
Additive Replacement Level (%)	11	7.9	27.90	53	48.9	7.64	23	24.6	7.17
Additive SSA (cm².g⁻¹)	1577	1676	6.30	15,502	16,744	8.01	21804.95	21,608	0.90
	18% Silica Fume			23% Silica Fume			13% Quartz		
	Actual	Predicted	Error (%)	Actual	Predicted	Error (%)	Actual	Predicted	Error (%)
Additive Type	4	4	0	4	4	0	1	1	0
Additive Replacement Level (%)	18	21.7	21.03	23	28.3	23.30	13	12.0	7.51
Additive SSA (cm².g⁻¹)	198,000	198,000	0	198,000	198,000	0	14,970	14,442	3.53

5. Conclusion

In the past decade or so, machine learning (ML) has arisen as a promising, holistic approach to reveal such composition-property correlations in composite materials. This study describes the use of a Random Forests (RF) model to enable high-fidelity predictions of time-dependent hydration kinetics of blended ordinary Portland cement (OPC) systems — more specifically [OPC + mineral additives] systems — using the systems' mix design features and physiochemical attributes as inputs. As a method to limit OPC's contributions to global anthropogenic CO₂ emissions, mineral additives have been endorsed by literature as partial replacements for OPC. However, it has been well established that mineral replacements of differing types and physical properties can have varying effects on the hydration kinetics of OPC-based systems. Therefore — in regards to more complex systems — it is infeasible for semi-empirical kinetic models to reveal the underlying nonlinear composition-property (i.e., reactivity) relationships.

To elucidate the aforementioned underlying nonlinear composition-property (i.e., reactivity) relationships, an original application of the Random Forests (RF) ML model was utilized to predict the time-dependent hydration behavior (i.e., heat flow rate; and cumulative heat release) of plain and blended ordinary Portland cement (OPC) systems. The prediction results have proven that the RF model has the ability to predict and optimize the relatively continuous (i.e., short time steps) and long time period (i.e., 24 hour) heat-evolution-determined kinetic profiles corresponding to plain and blended OPC systems. Further, these novel ML results demonstrate for the first time that rapid and reliable predictions — once the model is properly and rigorously trained, are possible for new OPC-based systems without further experimentation — of

time-dependent hydration behavior of plain and blended OPC systems are indeed feasible, a feat that is currently impossible with current numerical kinetic models. To the best of the authors' knowledge, this is the first study that employs ML to predict time-dependent kinetic behavior of cement hydration that features multiple mineral additives and focuses on mixture design parameters and physiochemical attributes.

The results from this study can be expanded to formulate mixture designs that satisfy target (user-imposed) kinetic criteria, even without a comprehensive understanding of the underlying kinetic mechanisms. For example, if the current training database is extended to include additional [OPC + mineral additives] permutations, the heat-evolution profiles corresponding to the hydration of more complex blended systems could be potentially predicted with reasonable accuracy. Such examples include, a system with numerous mineral additives, such as a [OPC + limestone + quartz + metakaolin + silica fume] system or even a system containing other commonly used additives such as fly ashes and/or geopolymers if said additives are included in additional permutations. In terms of optimization, a database with additional data-entries could also be utilized to predict, in a high-fidelity approach, the exact ideal amount and additive type for a [OPC + mineral additive(s)] system with a desired time-dependent heat-evolution profile and strength threshold. Of course, if other mineral additives are introduced, the database must be developed, accordingly. For example, it can be estimated that ≥ 100 additional binary/ternary [cement + fly ash + other mineral additives] data entries — featuring variations in fly ash composition, fineness, and replacement levels — should be added to the existing training database to enable reliable predictions of cement hydration behavior. As the database is grown

(i.e., database volume increases), the variations in physicochemical attributes will span gradually wider, and the prediction performance will improve, albeit at progressively slower rates.

Lastly, there are additionally a plethora of potential opportunities for improvement within the prediction and optimization processes in future studies; however, the novel work within this publication represents an important initial design optimization milestone for cementitious mixture designs that successfully demonstrates the feasible for a given ML model to elucidate underlying nonlinear time-dependent composition-property (i.e., reactivity) relationships.

CRediT authorship contribution statement

Rachel Cook: Investigation, Data curation, Writing - original draft, Writing - review & editing, Supervision. **Taihao Han:** Software, Writing - original draft, Writing - review & editing. **Alaina Childers:** Investigation. **Cambria Ryckman:** Investigation. **Kamal Khayat:** Conceptualization, Funding acquisition. **Hongyan Ma:** Conceptualization, Funding acquisition. **Jie Huang:** Conceptualization, Funding acquisition. **Aditya Kumar:** Conceptualization, Funding acquisition, Writing - review & editing, Supervision.

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.

Acknowledgements

The authors acknowledge financial support for this research provided by the UM system; the Federal Highway Administration (Award no: 693JJ31950021); the Leonard Wood Institute (LWI: W911NF-07-2-0062) and the National Science Foundation (NSF-CMMI: 1661609 and 1932690; and NSF-DMR: 2034856). The authors would like to gratefully thank Theodore Lewis and Missouri S&T for providing the assistance and facilities, respectively, to accomplish both the experimental and computational work portions of the research detailed in this communication.

Data availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.matdes.2021.109920>.

References

- [1] K.L. Scrivener, Straight talk with Karen Scrivener on cements, CO₂ and sustainable development, *Am. Ceram. Soc. Bull.* 91 (2012) 47–50.
- [2] K.C. Curry, Mineral Commodity Summaries 2020, U.S. Department of the Interior, Reston, Virginia, 2020.
- [3] E. Gartner, H. Hirao, A review of alternative approaches to the reduction of CO₂ emissions associated with the manufacture of the binder phase in concrete, *Cement Concr. Res.* 78(Part A) (2015) 126–142. <https://doi.org/10.1016/j.cemconres.2015.04.012>.
- [4] M. Schneider, Process technology for efficient and sustainable cement production, *Cement Concr. Res.* 78(Part A) (2015) 14–23. <https://doi.org/10.1016/j.cemconres.2015.05.014>.
- [5] H.-M. Ludwig, W. Zhang, Research review of cement clinker chemistry, *Cement Concr. Res.* 78(Part A) (2015) 24–37. <https://doi.org/10.1016/j.cemconres.2015.05.018>.
- [6] H.G. van Oss, USGS, Cement – Mineral Commodity Summaries, U.S. Department of the Interior, 2018.
- [7] J.J. Biernacki, J.W. Bullard, G. Sant, K. Brown, F.P. Glasser, S. Jones, T. Ley, R.A. Livingston, L. Nicoleau, J. Olek, F. Sanchez, R. Shahsavari, P.E. Stutzman, K. Soblev, T. Prater, Cements in the 21st century: Challenges, perspectives, and opportunities, 100 (2017) 27.
- [8] E. Gartner, T. Sui, Alternative cement clinkers, *Cem. Concr. Res.* 114 (2018) 27–39. <https://doi.org/10.1016/j.cemconres.2017.02.002>.
- [9] E. Worrell, L. Price, N. Martin, C. Hendriks, L.O. Meida, Carbon dioxide emissions from the global cement industry 1, *Annu. Rev. Energy Env.* 26 (2001) 303–329.
- [10] W.A. Gutteridge, J.A. Dalziel, Filler cement: the effect of the secondary component on the hydration of Portland cement: part I. A fine non-hydraulic filler, *Cement Concr. Res.* 20 (1990) 778–782.
- [11] W.A. Gutteridge, J.A. Dalziel, Filler cement: the effect of the secondary component on the hydration of Portland cement: part 2. fine hydraulic binders, *Cement Concr. Res.* 20 (1990) 853–861.
- [12] V. Rahhal, R. Talero, Early Hydration of Portland Cement With Crystalline Mineral Additions, *Cement Concr. Res.* 35 (2005) 1285–1291.
- [13] R.L. Sharma, S.P. Pandey, Influence of Mineral Additives on the Hydration Characteristics of Ordinary Portland Cement, *Cement Concr. Res.* 29 (1999) 1525–1529.
- [14] K. De Weerd, 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.
- [15] A.R. Jayapalan, B.Y. Lee, S.M. Fredrich, K.E. Kurtis, Influence of Additions of Anatase TiO₂ Nanoparticles on Early-Age Properties of Cement-Based Materials, *Transp. Res. Rec.* 2141 (2010) 41–46.
- [16] J.J. Thomas, H.M. Jennings, J.J. Chen, Influence of nucleation seeding on the hydration mechanisms of tricalcium silicate and cement, *J. Phys. Chem. C* 113 (2009) 4327–4334. <https://doi.org/10.1021/jp809811w>.
- [17] M. Bellotto, A. Gualtieri, G. Artioli, S.M. Clark, Kinetic Study of the Kaolinite-Mullite Reaction Sequence. Part I: Kaolinite Dehydroxylation, *Phys. Chem. Min.* 22 (1995) 207–217. <https://doi.org/10.1007/BF00202253>.
- [18] 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>.
- [19] A. Kumar, T. Oey, G. Falzone, J. Huang, M. Bauchy, M. Balonis, N. Neithalath, J. Bullard, G. Sant, The filler effect: The influence of filler content and type on the hydration rate of tricalcium silicate, *J. Am. Ceram. Soc.* 100 (2017) 3316–3328.
- [20] 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.
- [21] 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 (2019) 2764–2779. <https://doi.org/10.1111/jace.16936>.
- [22] 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.
- [23] 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.
- [24] A.M. Ley-Hernandez, J. Lapeyre, R. Cook, A. Kumar, D. Feys, Elucidating the Effect of Water Content on Hydration Mechanisms of Cement, *ACS Omega* 3 (2018) 5092–5105. <https://doi.org/10.1021/acsomega.8b00097>.
- [25] A. Ouzia, K. Scrivener, The needle model: A new model for the main hydration peak of alite, *Cement Concr. Res.* 115 (2019) 339–360.
- [26] K.L. Scrivener, P. Juilland, P.J. Monteiro, Advances in understanding hydration of Portland cement, *Cem. Concr. Res.* 78 (2015) 38–56.
- [27] J.J. Thomas, J.J. Biernacki, J.W. Bullard, S. Bishnoi, J.S. Dolado, G.W. Scherer, A. Luttge, Modeling and simulation of cement hydration kinetics and microstructure development, *Cem. Concr. Res.* 41 (2011) 1257–1278.
- [28] W.A. Johnson, R.F. Mehl, Reaction Kinetics in Processes of Nucleation and Growth, *Transactions of the American Institute of Mining.* 195 (1939) 416–442.
- [29] M. Avrami, Kinetics of phase change, I General theory, *The Journal of Chemical Physics.* 7 (1939) 1103–1112.
- [30] M. Avrami, Kinetics of Phase Change. II Transformation-Time Relations for Random Distribution of Nuclei, *J. Chem. Phys.* 8 (1940) 212–224.
- [31] M. Avrami, Granulation, phase change, and microstructure kinetics of phase change. III, *J. Chem. Phys.* 9 (1941) 177–184.
- [32] A.N. Kolmogorov, On the statistical theory of the crystallization of metals, *Bull. Acad. Sci. USSR, Math. Ser.* 1 (1937) 355–359.
- [33] A.J. Allen, J.C. McLaughlin, D.A. Neumann, R.A. Livingston, In situ quasi-elastic scattering characterization of particle size effects on the hydration of tricalcium silicate, 2004.
- [34] P.W. Brown, J. Pommersheim, G. Frohnsdorff, A Kinetic Model for the Hydration of Tricalcium Silicate, *Cement Concr. Res.* 15 (1985) 35–41.
- [35] A. Damasceni, L. Dei, E. Fratini, F. Ridi, S.-H. Chen, P. Baglioni, A Novel Approach Based on Differential Scanning Calorimetry Applied to the Study of Tricalcium Silicate Hydration Kinetics, *J. Phys. Chem. B* 106 (2002) 11572–11578. <https://doi.org/10.1021/jp020211>.
- [36] S.A. FitzGerald, D.A. Neumann, J.J. Rush, D.P. Bentz, R.A. Livingston, In situ quasi-elastic neutron scattering study of the hydration of tricalcium silicate, *Chem. Mater.* 10 (1998) 397–402.

- [37] S.A. Grant, G.E. Boitnott, C.J. Korhonen, R.S. Sletten, Effect of temperature on hydration kinetics and polymerization of tricalcium silicate in stirred suspensions of CaO-saturated solutions, 36 (2006) 672–677. <https://doi.org/10.1016/j.cemconres.2005.10.001>.
- [38] N. Tenoutasse, A. De Donder, The kinetics and mechanism of hydration of tricalcium silicate, *Silicates Ind.* 35 (1970) 301–307.
- [39] J.J. Thomas, H.M. Jennings, Effects of D₂O and mixing on the early hydration kinetics of tricalcium silicate, *Chem. Mater.* 11 (1999) 1907–1914.
- [40] D.R. Vollet, A.F. Craievich, Effects of temperature and of the addition of accelerating and retarding agents on the kinetics of hydration of tricalcium silicate, *J. Phys. Chem. B* 104 (2000) 12143–12148.
- [41] I.G. Richardson, Tobermorite/jennite- and tobermorite/calcium hydroxide-based models for the structure of C-S-H: applicability to hardened pastes of tricalcium silicate, β -dicalcium silicate, Portland cement, and blends of Portland cement with blast-furnace slag, metakaolin, or silica fume, *Cement Concr. Res.* 34 (2004) 1733–1777. <https://doi.org/10.1016/j.cemconres.2004.05.034>.
- [42] A. Nonat, The structure and stoichiometry of C-S-H, *Cem. Concr. Res.* 34 (2004) 1521–1528. <https://doi.org/10.1016/j.cemconres.2004.04.035>.
- [43] A.J. Allen, R.C. Oberthur, D. Pearson, P. Schofield, C.R. Wilding, Development of the fine porosity and gel structure of hydrating cement systems, *Philos. Mag.* 56 (1987) 263–288. <https://doi.org/10.1080/13642818708221317>.
- [44] J.J. Thomas, A new approach to modeling the nucleation and growth kinetics of tricalcium silicate hydration, *J. Am. Ceram. Soc.* 90 (2007) 3282–3288.
- [45] J.W. Cahn, The kinetics of grain boundary nucleated reactions, *Acta Metall.* 4 (1956) 449–459.
- [46] J.W. Bullard, A determination of hydration mechanisms for tricalcium silicate using a kinetic cellular automaton model, *J. Am. Ceram. Soc.* 91 (2008) 2088–2097.
- [47] J.W. Bullard, A three-dimensional microstructural model of reactions and transport in aqueous mineral systems, *Modell. Simul. Mater. Sci. Eng.* 15 (2007) 711.
- [48] F. Ridi, E. Fratini, P. Luciani, F. Winnefeld, P. Baglioni, Tricalcium silicate hydration reaction in the presence of comb-shaped superplasticizers: boundary nucleation and growth model applied to polymer-modified pastes, *J. Phys. Chem. C* 116 (2012) 10887–10895.
- [49] L. Valentini, M. Favero, M.C. Dalconi, V. Russo, G. Ferrari, G. Artioli, Kinetic Model of Calcium-Silicate Hydrate Nucleation and Growth in the Presence of PCE Superplasticizers, *Cryst. Growth Des.* (2016) 646–654.
- [50] A. Kumar, S. Bishnoi, K.L. Scrivener, Modelling early age hydration kinetics of alite, *Cem. Concr. Res.* 42 (2012) 903–918.
- [51] V.K. Peterson, M.C.G. Juenger, Hydration of tricalcium silicate: effects of CaCl₂ and sucrose on reaction kinetics and product formation, *Chem. Mater.* 18 (2006) 5798–5804.
- [52] V.K. Peterson, A.E. Whitten, Hydration processes in tricalcium silicate: application of the boundary nucleation model to quasielastic neutron scattering data, *J. Phys. Chem. C* 113 (2009) 2347–2351.
- [53] G.W. Scherer, J. Zhang, J.J. Thomas, Nucleation and growth models for hydration of cement, *Cem. Concr. Res.* 42 (2012) 982–993.
- [54] S. Garraut, T. Behr, A. Nonat, Formation of the C-S-H Layer During Early Hydration of Tricalcium Silicate Grains with Different Sizes, *J. Phys. Chem. B* 110 (2006) 270–275.
- [55] E. Masoero, J.J. Thomas, H.M. Jennings, A Reaction Zone Hypothesis for the Effects of Particle Size and Water-to-Cement Ratio on the Early Hydration Kinetics of C₃S, *J. Am. Ceram. Soc.* 97 (2014) 967–975.
- [56] R. Cook, H. Ma, A. Kumar, Mechanism of tricalcium silicate hydration in the presence of polycarboxylate polymers, *SN Appl. Sci.* 1 (2019) 145. <https://doi.org/10.1007/s42452-018-0153-1>.
- [57] J. Lapeyre, H. Ma, A. Kumar, Effect of particle size distribution of metakaolin on hydration kinetics of tricalcium silicate, 102 (2019) 5976–5988. <https://doi.org/10.1111/jace.16467>.
- [58] C. Naber, F. Bellmann, T. Sowoidnich, F. Goetz-Neunhoeffer, J. Neubauer, Alite dissolution and C-S-H precipitation rates during hydration, *Cement Concr. Res.* 115 (2019) 283–293. <https://doi.org/10.1016/j.cemconres.2018.09.001>.
- [59] L. Nicoleau, A. Nonat, A new view on the kinetics of tricalcium silicate hydration, *Cement Concr. Res.* 86 (2016) 1–11. <https://doi.org/10.1016/j.cemconres.2016.04.009>.
- [60] 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. Civil Eng.* 31 (2019) 04019255. [https://doi.org/10.1061/\(ASCE\)JMT.1943-5533.0002902](https://doi.org/10.1061/(ASCE)JMT.1943-5533.0002902).
- [61] 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). <https://doi.org/10.1016/j.conbuildmat.2020.118271>.
- [62] J.-S. Chou, C.-F. Tsai, A.-D. Pham, Y.-H. Lu, Machine learning in concrete strength simulations: Multi-nation data analytics, *Constr. Build. Mater.* 73 (2014) 771–780. <https://doi.org/10.1016/j.conbuildmat.2014.09.054>.
- [63] A. Behnood, V. Behnood, M.M. Gharehveran, K.E. Alyamac, Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm, *Constr. Build. Mater.* 142 (2017) 199–207. <https://doi.org/10.1016/j.conbuildmat.2017.03.061>.
- [64] J.-S. Chou, C.-K. Chiu, M. Farfoura, I. Al-Taharwa, Optimizing the prediction accuracy of concrete compressive strength based on a comparison of data-mining techniques, *J. Comput. Civil Eng.* 25 (2010) 242–253.
- [65] B.A. Omran, Q. Chen, R. Jin, Comparison of data mining techniques for predicting compressive strength of environmentally friendly concrete, *J. Comput. Civil Eng.* 30 (2016) 04016029.
- [66] I.-C. Yeh, Modeling of strength of high-performance concrete using artificial neural networks, *Cem. Concr. Res.* 28 (1998) 1797–1808.
- [67] Z.H. Duan, S.C. Kou, C.S. Poon, Using artificial neural networks for predicting the elastic modulus of recycled aggregate concrete, *Constr. Build. Mater.* 44 (2013) 524–532. <https://doi.org/10.1016/j.conbuildmat.2013.02.064>.
- [68] D. Cruz, D.A. Talbert, W. Eberle, J. Biernacki, A Neural Network Approach for Predicting Microstructure Development in Cement, 2016. <https://www.semanticscholar.org/paper/A-Neural-Network-Approach-for-Predicting-in-Cement-Cruz-Talbert/0306a66e8574e39cf4700f83d0c6ef0606b4540f> (accessed April 7, 2020).
- [69] S.S. Bangaru, C. Wang, M. Hassan, H.W. Jeon, T. Ayiluri, Estimation of the degree of hydration of concrete through automated machine learning based microstructure analysis – A study on effect of image magnification, *Adv. Eng. Inf.* 42 (2019). <https://doi.org/10.1016/j.aei.2019.100975>.
- [70] K.O. Akande, T.O. Owolabi, S. Twaha, S.O. Olatunji, Performance comparison of SVM and ANN in predicting compressive strength of concrete, *IOSR J. Comput. Eng.* 16 (2014) 88–94.
- [71] Z.-H. Duan, S.-C. Kou, C.-S. Poon, Prediction of compressive strength of recycled aggregate concrete using artificial neural networks, *Constr. Build. Mater.* 40 (2013) 1200–1206.
- [72] R. Gupta, M.A. Kewalramani, A. Goel, Prediction of concrete strength using neural-expert system, *J. Mater. Civ. Eng.* 18 (2006) 462–466.
- [73] J. Kasperkiewicz, J. Racz, A. Dubrawski, HPC strength prediction using artificial neural network, *J. Comput. Civil Eng.* 9 (1995) 279–284.
- [74] N.K. Nagwani, S.V. Deo, Estimating the concrete compressive strength using hard clustering and fuzzy clustering based regression techniques, *Sci. World J.* 2014 (2014).
- [75] V. Veloso de Melo, W. Banzhaf, Improving the prediction of material properties of concrete using Kaizen Programming with Simulated Annealing, *Neurocomputing*. 246 (2017) 25–44. <https://doi.org/10.1016/j.neucom.2016.12.077>.
- [76] I.-C. Yeh, Modeling concrete strength with augment-neuron networks, *J. Mater. Civ. Eng.* 10 (1998) 263–268.
- [77] I.-C. Yeh, L.-C. Lien, Knowledge discovery of concrete material using genetic operation trees, *Expert Syst. Appl.* 36 (2009) 5807–5812.
- [78] B.A. Young, A. Hall, L. Pilon, P. Gupta, G. Sant, Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods, *Cem. Concr. Res.* 115 (2019) 379–388.
- [79] M.F. Zarandi, I.B. Türksen, J. Sobhani, A.A. Ramezani-pour, Fuzzy polynomial neural networks for approximation of the compressive strength of concrete, *Appl. Soft Comput.* 8 (2008) 488–498.
- [80] M. Antoni, J. Rossen, F. Martirena, K. Scrivener, Cement substitution by a combination of metakaolin and limestone, *Cem. Concr. Res.* 42 (2012) 1579–1589.
- [81] K. Vance, M. Aguayo, T. Oey, G. Sant, N. Neithalath, Hydration and strength development in ternary portland cement blends containing limestone and fly ash or metakaolin, *Cem. Concr. Compos.* 39 (2013) 93–103.
- [82] A.M. Ramezani-pour, R.D. Hooten, A study on hydration, compressive strength, and porosity of Portland-limestone cement mixes containing SCMs, *Cem. Concr. Compos.* 51 (2014) 1–13.
- [83] M. Boháč, M. Palou, R. Novotný, J. Másilko, D. Všianský, T. Staněka, Investigation on early hydration of ternary Portland cement–blast–furnace slag–metakaolin blends, *Constr. Build. Mater.* 64 (2014) 333–341.
- [84] T. Matschei, B. Lothenbach, F.P. Glasser, The role of calcium carbonate in cement hydration, *Cem. Concr. Res.* 37 (2007) 551–558.
- [85] 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>.
- [86] R. Cook, C.M. Keitumetse, M.B. Hayat, A. Kumar, L. Alagha, Prediction of Flotation Performance of Sulfide Minerals Using an Original Hybrid Machine Learning Model Article in Press, *Eng. Rep.* (2020) 1–33. <https://doi.org/10.1002/eng2.12167>.
- [87] Y. Zhuang, Q. Yang, T. Han, R. O'Malley, A. Kumar, R.E. Gerald, J. Huang, Fiber optic sensor embedded smart helmet for real-time impact sensing and analysis through machine learning, *J. Neurosci. Methods* 351 (2021). <https://doi.org/10.1016/j.jneumeth.2021.109073>.
- [88] 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). <https://doi.org/10.1016/j.cemconcomp.2020.103863>.
- [89] 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). <https://doi.org/10.1016/j.cemconres.2020.106164>.
- [90] L. Breiman, Bagging predictors, *Mach. Learn.* 24 (1996) 123–140. <https://doi.org/10.1007/BF00058655>.
- [91] L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32.
- [92] X. Chen, H. Ishwaran, Random forests for genomic data analysis, *Genomics* 99 (2012) 323–329.
- [93] G.Š. Biau, L. Devroye, G.Š. Lugosi, Consistency of random forests and other averaging classifiers, *J. Mach. Learn. Res.* 9 (2008) 2015–2033.

- [94] C. Schaffer, Selecting a classification method by cross-validation, *Mach. Learn.* 13 (1993) 135–143, <https://doi.org/10.1007/BF00993106>.
- [95] T.G. Dietterich, Ensemble methods in machine learning, in: *International Workshop on Multiple Classifier Systems*, 2000, pp. 1–15, https://doi.org/10.1007/3-540-45014-9_1.
- [96] T. Eitrich, B. Lang, Efficient optimization of support vector machine learning parameters for unbalanced datasets, *J. Comput. Appl. Math.* 196 (2006) 425–436, <https://doi.org/10.1016/j.cam.2005.09.009>.
- [97] J. Bergstra, Y. Bengio, Random Search for Hyper-Parameter Optimization, *J. Mach. Learn. Res.* 13 (2012) 281–305.
- [98] M.R. Segal, Machine Learning Benchmarks and Random Forest Regression, 2004. <https://escholarship.org/uc/item/35x3v9t4> (accessed April 18, 2020).
- [99] N.M. Anoop Krishnan, S. Mangalathu, M.M. Smedskjaer, A. Tandia, H. Burton, M. Bauchy, Predicting the dissolution kinetics of silicate glasses using machine learning, *J. Non-Cryst. Solids* 487 (2018) 37–45, <https://doi.org/10.1016/j.jnoncrysol.2018.02.023>.
- [100] V. Svetnik, A. Liaw, C. Tong, J.C. Culberson, R.P. Sheridan, B.P. Feuston, Random forest: a classification and regression tool for compound classification and QSAR modeling, *J. Chem. Inf. Comput. Sci.* 43 (2003) 1947–1958.
- [101] M. Pelikan, Hierarchical Bayesian Optimization Algorithm, in: M. Pelikan (Ed.), *Hierarchical Bayesian Optimization Algorithm: Toward a New Generation of Evolutionary Algorithms*, Springer, Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 105–129, https://doi.org/10.1007/978-3-540-32373-0_6.
- [102] K. Swersky, J. Snoek, R.P. Adams, Multi-Task Bayesian Optimization, in: C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K.Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 26*, Curran Associates, Inc., 2013, pp. 2004–2012. <http://papers.nips.cc/paper/5086-multi-task-bayesian-optimization.pdf> (accessed June 19, 2018).