



Sequential learning to accelerate discovery of alkali-activated binders

Christoph Völker^{1,*} , Rafia Firdous², Dietmar Stephan², and Sabine Kruschwitz^{2,3}

¹Bundesanstalt für Materialforschung und -prüfung: Bundesanstalt für Materialforschung und -prüfung, Berlin, Germany

²Technical University of Berlin, Berlin, Germany

³Federal Institute for Materials Research and Testing, Berlin, Germany

Received: 27 April 2021

Accepted: 5 July 2021

Published online:

19 July 2021

© The Author(s) 2021

ABSTRACT

Alkali-activated binders (AAB) can provide a clean alternative to conventional cement in terms of CO₂ emissions. However, as yet there are no sufficiently accurate material models to effectively predict the AAB properties, thus making optimal mix design highly costly and reducing the attractiveness of such binders. This work adopts sequential learning (SL) in high-dimensional material spaces (consisting of composition and processing data) to find AABs that exhibit desired properties. The SL approach combines machine learning models and feedback from real experiments. For this purpose, 131 data points were collected from different publications. The data sources are described in detail, and the differences between the binders are discussed. The sought-after target property is the compressive strength of the binders after 28 days. The success is benchmarked in terms of the number of experiments required to find materials with the desired strength. The influence of some constraints was systematically analyzed, e.g., the possibility to parallelize the experiments, the influence of the chosen algorithm and the size of the training data set. The results show the advantage of SL, i.e., the amount of data required can potentially be reduced by at least one order of magnitude compared to traditional machine learning models, while at the same time exploiting highly complex information. This brings applications in laboratory practice within reach.

Handling Editor: Maude Jimenez.

Address correspondence to E-mail: christoph.voelker@bam.de

<https://doi.org/10.1007/s10853-021-06324-z>

Introduction

Concrete is the most widely used building material on earth. In 2019 alone, worldwide 30 billion tons of concrete was produced—that is almost four tons for every single person [1]. Since the cement is the most important constituent of concrete and its production is associated with the emission of CO₂, around eight percent of man-made CO₂ emissions come from the production of cement [2]. The increasing demand for infrastructure development throughout the globe related to increasing population and consequently economic growth leads to continued high consumption of cement-based building materials. If the climate agreements are adhered to, the question arises as to what extent ordinary Portland cement (OPC) as a building material is still viable as a mass product [3]. In this regard, the cement technology roadmap defined several ways to reduce the CO₂ footprint at every step of cement production and huge effort has been made to develop more environmentally friendly binders [4]. However, exploring more CO₂-friendly alternatives requires an extensive experimental program and, consequently, long development cycles until they can be used in practice [5]. Alkali-activated binders (AAB) and/or geopolymers (GP) omit the energy-intensive kiln process that is responsible for a large part of the CO₂ emissions of conventional types of cement and therefore are considered as environmentally friendly alternatives [6].

AAB/GP are synthesized by the reaction of an aluminosilicate precursor with an alkaline source [7]. Natural or artificial pozzolans, including granulated blast furnace slag, fly ash, metakaolin, or natural pozzolans, can be used as aluminosilicate sources [7]. However, the nature, chemical and mineralogical composition of the aluminosilicate precursor affect the properties of the resultant binder, such as fresh and hardened characteristics [8]. On the one hand, the wide variety of precursors available for the synthesis of AAB/GP makes them a suitable alternative binder for many applications. On the other hand, such large variations in the precursor composition make it difficult to draw general valid rules to synthesize these binders with the expected properties. Moreover, the type and concentration of alkaline solution impact the reactivity of the aluminosilicate precursors used for AAB/GP [8]. Several publications have shown the potential of granulated blast

furnace slag and fly ash as a precursor for AAB/GP [5]. However, the environmental protection laws in many countries and European environmental goals aim to reduce and eventually omit the burning of coal and encourage renewable energies [9]. The German environmental goals target the closing of coal power plants by the year 2038 [10], which means the available amount of fly ash from local sources in Germany will decrease in some years. Similarly, the iron industry aims at using alternative renewable fuel sources for its manufacturing process, which means that the composition and quality and probably also the amount of blast furnace slag will be affected [11]. Moreover, for more than a decade, fly ash, bottom ash and boiler slags are fully utilized in European countries in either producing blended cements, direct addition to concrete or other applications [5, 12]. Henceforth, the decreasing amount of common pozzolans and their full utilization by industry lead to the thrive of searching for new alternative precursors, including natural (such as volcanic ashes) and artificial resources (such as urban and agro-industrial wastes) [13]. The increasing demand for the search for alternative aluminosilicate sources and the complexity of the synthesis of AAB/GP binders makes it challenging to gain fast progress in this field.

Objective, scope and novelty of the research

In order to achieve the desired breakthrough in the shortest possible time, optimization of the materials research process is urgently needed. Sequential Learning (SL) and the closely related Bayesian optimization have repeatedly been reported to have great potential in accelerating drug and material discovery [14, 15]. The basic idea is to reduce the number of unsuccessful experiments (i.e., that lead to materials with unwanted properties) so that an ideal sequence of successive experiments is achieved. This is accomplished by coupling a prediction model (e.g., a machine learning model) with a decision-making rule based on a so-called utility function that guides the experimental program. Work that argued for the fundamental predictability of the properties of cementitious materials in general and AABs in particular with data models has been published (see Table 1). However, to the best of the authors' knowledge, there is no work available that investigates the transferability of SL methods to AAB research.

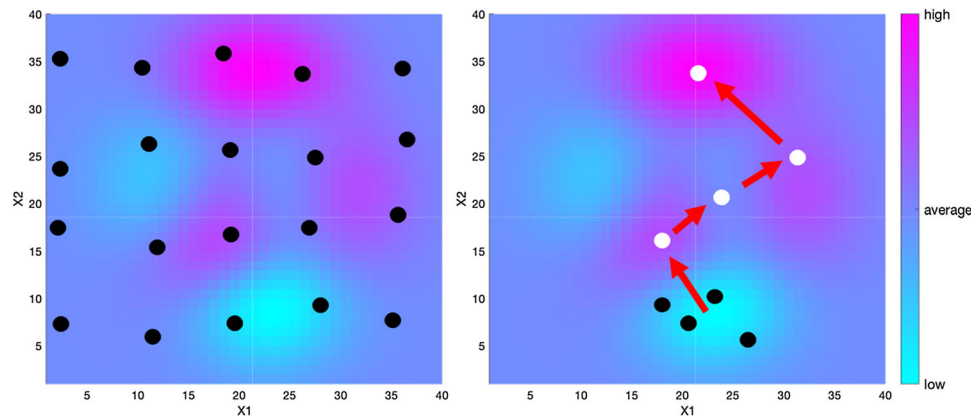


Figure 1 Schematic of sampling for discovering high performance materials (pink areas) with data-based methods; Left: (nearly) identically distributed sampling (black dots) as

aspired in classic Machine Learning; right: sampling with Sequential Learning with initial data set (black dots) and sequential sampling (red arrows) of new samples (white dots).

Figure 1 illustrates the conceptual differences between classic Machine Learning (ML) (Fig. 1, left) and the novel SL approach (Fig. 1, right). The figure shows a mathematical space spanned by two base materials X_1 and by X_2 (in the real scenario, additional features maybe considered). Depending on the respective mixing ratio, the synthesis of X_1 and X_2 results in materials whose resulting material properties are shown in color (turquoise areas for low performance and pink areas for high performance materials).

ML and SL model this relationship using sampled data (black and white dots in Fig. 1, respectively). These models can be employed to predict ideal combinations of X_1 and X_2 . Classical ML approaches accomplish this (mostly) through interpolation, whereas SL is inherently extrapolative and thus potentially requires much fewer sample data. However, serial data collection of SL, even if successful, may be disadvantageous, because waiting for experimental results could delay experimental progress. This is especially the case for materials whose synthesis is complex and whose material properties require time to develop or characterize (e.g., 28-day compressive strength of binders). Collecting all samples at once or in batches can be more successful. SL could therefore fill a gap in material innovation where data are not available or large-scale data collection would be too expensive.

The scope of this paper is to investigate exactly under which conditions SL can contribute to accelerating research regarding the properties of alkali-

activated binders. For this, we have compiled an AAB data set with 131 samples from several publications. Based on these data, we investigate how the SL approach, the complexity of the task, the quantitative research objective and the desired success rate influence the performance of SL and deduce some of the critical circumstances under which SL can potentially enhance AAB research practice. We introduce a novel utility function that adapts common utility functions for applications with minimal training data, i.e., lower number of experiments to reach the optimal sample design.

Structure of the research

The presented paper is organized as follows. First, we survey the literature in ML for concrete property prediction in terms of its potential and applicability for alkali-activated binder research. We combine this with an overview of current AAB research to highlight the practical potential associated with an optimized search by SL. We then briefly describe the SL algorithms which were used in this work. We use Gaussian processing regression (GPR), decision tree regression (DT) and tree ensemble regression (TE) and four different utility functions, namely maximum expected improvement (MEI), maximum likelihood of improvement (MLI) and introduce a novel combination of MEI & MLI with a maximized distance measure, respectively. We go on to describe how we benchmark the acceleration in the experimental program. In the experimental section, we

explain how data were acquired from literature and describe our investigations with SL. The result section shows which constellations lead to a successful acceleration of the experimental progress. We conclude by discussing how scientists can integrate SL into their workflow to fully exploit the potential of SL.

Theoretical background and methods

Literature review

Because of various properties and many possible combinations of the constituents of building materials, early on, a computer-aided approach was considered [16]. Beyond the mere streamlining of the concrete mix design, capabilities have expanded significantly by introducing ML approaches. Models that interpolate the properties of new mixtures are generated from data of existing compositions where the properties have been determined in the laboratory. Naderpour et al. [17] summarized some of the earlier models. Chaabene et al. [18] review the available literature on ML for the prediction of mechanical properties of concrete. Table 1 focuses on recent research in this field by summarizing models from the past five years (2017–2021).

ML-based approaches exist for a wide range of cementitious materials, from OPC-based binders to AAB and recycled aggregate concrete, to name a few. In most cases, the models are used to predict compressive strength—one of the most important properties of concrete. With an average of 655 samples, the ML models require large amounts of data collected by laboratory standards. Remarkably, only 6.67 features (factors effecting the properties) are used on average to represent the composition of the material. The level of detail of the data considered is thus low relative to the complexity of alternative binders, as the properties of AAB are depended on several factors and the wide characteristics of aluminosilicate precursors and alkaline activators directly impact the properties of resultant AAB [8].

Based on the analysis of 111 distributed data sources, Xie et al. [46] state that much more detailed data are needed for a general understanding that goes beyond the limits of a single available AAB. This is a common challenge in materials science, where millions of possible compounds span a high-

dimensional discovery or search space, of which only a tiny fraction has been experimentally explored. The task of material discovery is to find desired properties in this space. The challenge is to make the best use of the limited knowledge from the few data points available. Currently, applied ML-based cement models are infeasible for this task because they cover only a few dimensions of the search space and require large amounts of data.

Despite success in other experimental sciences such as drug discovery, relatively few publications of SL exist in materials science. Lookman et al. [14] give an overview of the SL landscape in materials science. Despite the fact that SL regularly starts with a machine learning model based on only a few very high-dimensional data points and requires relatively little additional data from the laboratory, it often outperforms baseline benchmarks. However, these benchmarks are mostly statistically reached and do not necessarily mean that SL will also accelerate research in practice. Here, experimental designs often determine the speed and SL would possibly entail an adjustment of the entire process. Yet even in simple statistically motivated scenarios, specific performance is usually highly dependent on the data and the problem and the exact relationships are still largely unknown in [47]. Lookman et al. conclude that despite many studies that use machine learning to make predictions, feedback between ML and experiments via SL has only recently been investigated in materials science.

The following section surveys what information is required to be able to make predictions of AAB properties. The available literature about AAB/GP shows that the properties of these binders, such as compressive strength, depend on a vast variety of factors, far beyond those considered in Table 1. These factors may include the chemical and mineralogical composition of the precursor, type, concentration, $\text{SiO}_2/\text{M}_2\text{O}$ (M = alkali cation) of the alkaline solution used, curing conditions, specific surface area, water-to-binder ratio, degree of silicate polymerization, age of the sample and others [8, 48–50]. The aluminosilicate source materials for alkali-activation show a wide range of oxide composition and mineralogical phases. The chemistry of the reactive phase mostly defines the reactivity of the raw material. However, the composition of the alkaline solution has an enormous impact on the strength developing phases [8, 50]. An optimum range of silica modulus is

Table 1 Summary of recently published Machine Learning models for the prediction of concrete properties

Reference	Material	ML-Model	Target	Features	Samples
[19]	Alkali-activated binders	Fly ash-based concrete, alkali-activated slag concrete	Stepwise regression	Compressive strength	5 5168 (646 unique batches)
[20]		Alkali-activated cement (Boroaluminosilicate binder)	Artificial neural networks	Compressive strength	5 114
[21]		Alkali-activated cement	Random forest regression	Compressive strength, slump flow	20 180
[22]		Fly ash-based concrete	Decision trees, Bagged tree Ensembles, Gene expression programming	Compressive strength	8 270
[23]		Fly ash-based cement	(Hybrid) artificial neural networks	Compressive strength	4 210
[17]	Recycled aggregates	Recycled aggregate concrete	Artificial neural networks	Compressive strength	7 139
[24]		Recycled aggregate concrete	Artificial neural networks	Compressive strength	9 234
[25]		Recycled aggregate concrete	Artificial neural networks	Compressive strength	2 74
[26]		Recycled aggregate concrete	Artificial neural networks	tensile strength	4 421
[27]		Recycled aggregate concrete	Artificial neural networks	elastic modulus	7 400
[28]		Concrete containing waste foundry sand	M5P-tree	Compressive strength, elastic modulus, tensile strength	7/7/7 470/172/295
[29]		Recycled aggregate concrete	M5-tree, support vector regression	Compressive strength, elastic modulus, tensile strength	5/5/5 650/421/346
[30]		High-performance concrete	Support vector machine	Compressive strength	9 1761
[31]	Ordinary Portland cement	OPC with fly ash supplements	Artificial neural networks	Compressive strength	6 114
[32]		OPC	Artificial neural networks	Compressive strength, slump	7/7 482/295
[33]		OPC	Artificial neural networks	Concrete mix design (cement, water, fine & coarse aggregates) for a desired level of compressive strength	4 662
[34]		High-performance concrete, OPC	(Hybrid) artificial neural networks	Compressive strength	8 2817
[35]		High-performance concrete, OPC	M5P-tree	Compressive strength	6 1912
[36]		High-performance concrete, OPC	Random forest	Compressive strength	8 1030
[37]		OPC	Adaptive boosting	compressive strength	8 1030
[38]		OPC	Support vector machines, high-order response surface method	Compressive strength	5 Unknown

Table 1 continued

Reference	Material	ML-Model	Target	Features	Samples
[39]	Cementitious multi-component materials	random forest regression	Time-dependent hydration kinetics (rate and extent of cement hydration in the first 24 h after mixing)	7	249
[40]	Nano-TiO ₂ -modified mortar	Artificial neural networks	Compressive strength	5	194
[41]	Other Lightweight self-compacting concrete	Random forest regression	Compressive strength	8	131
[42]	Lightweight concrete	M5-tree, support vector regression, multivariate adaptive regression spline, decision tree regression	Compressive strength	4	91
[43]	Self-compacting concrete	Artificial neural networks	Compressive strength	11	205
[44]	Rubberized concrete	Artificial neural networks	Compressive strength	4	324
[45]	Rubberized concrete	Artificial neural networks	Compressive strength	6	112
Average				6.67	655

necessary to achieve higher compressive strength for natural pozzolan-based geopolymers while the other effecting factors were kept the same [8]. This optimum range of silica modulus, in turn, changed for every type of raw material. For various precursors, different alkaline solutions in different concentrations have been recommended to achieve the desired properties [49].

Similarly, the specific surface area measured either in terms of Blaine fineness or particle size distribution (d_{90} , d_{50} , d_{10}) affects the degree of reaction and eventually the compressive strength [7]. The amount of water present in these binders has a complex role and acts as a reaction medium. The extend of silicate polymerization of silicate solution influences its reactivity. Consequently, the number of experiments to be conducted to reach the optimum compressive strength can be extremely high and the certainty of reaching the optimum is low. Given this complex material behavior and the amount of experimentation conventionally required, advanced tools like SL have enormous potential to accelerate the material discovery process.

Setting up sequential learning for materials discovery

The underlying idea of SL is that not all experiments are equally useful. Some experiments provide more information than others. In contrast with classical design of experiments, where (only) the experimental

parameters are optimized, the potential outcomes of the experiments themselves are the decisive factor. The most promising experiments are preferred over dead-end experiments and experiments whose outcome is already known. Experimental results are used to iteratively improve the ML model with high-quality data. Each new experiment is selected to maximize the amount of useful information, e.g., according to [51], using previous experiments as a guide for the next experiment.

Figure 2 provides a chart that depicts the workflow of SL. The prediction of material properties in SL is based on a list of candidate materials given by experts using their domain knowledge, see (1) in Fig. 2. Materials may be of interest because they are available, cheap, known to have further desirable properties, or simply because they seem generally promising. Although the exact criteria are not specified, it is recognized that the performance of the SL for material discovery is related to the quality of the candidates [47]. The candidate materials are represented in the so-called design space (DS)—a vector space that is comparable to the feature space in classical ML approaches. In the DS, the coordinates of each material are parameterized information about raw material, (micro-) structure and processing. An initial training data set with known target properties serves as an input for the prediction model in the first round (see (2) in Fig. 2).

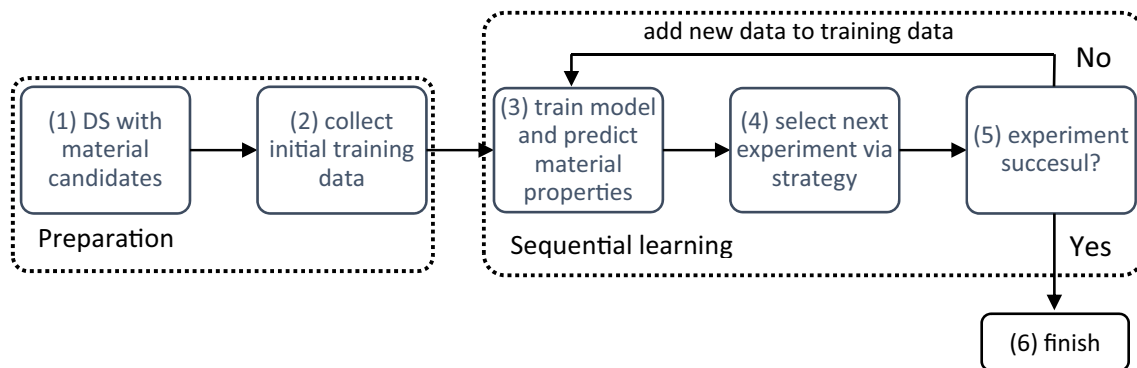


Figure 2 Workflow of Sequential Learning for materials discovery.

At the core of the iterative SL task is the prediction of experimental outcomes ((3) in Fig. 2), weighting the expected utility and deciding which candidate to investigate next ((4) in Fig. 2). The utility is commonly estimated based on the predicted material property (the closer a predicted experimental result is to the desired value, the more useful it is) and a measure of uncertainty. The latter is a key driver for discovering new relationships and the basis of experimenting in general. As Reyes et al. [15] aptly summarize it, “Actually, the outcome of an experiment is the deviation from what we expected.” In other words, if the outcome of an experiment is already known, there is no reason to conduct it and an experiment can be more useful if the uncertainty of its outcome is large. Lookman et al. [14] even state that they are not aware of an SL study where a new material has been discovered without utilizing uncertainties. In this sense, uncertainty can be considered an essential factor in the decision-making process. The SL task is finished as soon as the desired property is obtained, (5) and (6) in Fig. 2.

The following section lays out the prediction methods, uncertainty estimates and selection strategies utilized in this work.

Prediction methods and uncertainty estimates

We compared common regression methods against Gaussian Process Regression (GPR), which has its origin in adaptive sampling and is still one of the reference methods in SL or Bayesian optimization [14]. The regression methods investigated include shallow neural network regression (NNR), linear regression (LR), decision tree regression (DT) and bagged decision tree ensembles (TE). Our

preliminary analysis showed that LR and NNR perform worse compared to others and that NNR additionally required much higher computational capacity. The drop in performance is probably due to the higher proneness of these methods to co-linearity in sparse DS. This effect could be minimized by regularization but would require additional hyperparameter tuning, further increasing the computational effort. Based on these observations, we concentrate on the comparison of GPR, DT and TE in this paper. These methods are briefly described as follows.

Originally, decision trees and tree ensembles are classification algorithms that learn the segmentation of an input data space, e.g., the DS, from pairs of data and labels [52]. By introducing one class per discrete label value (and interpolated intermediate values), pseudo-regression is performed, meaning that interpolated predictions are possible, but extrapolations outside the range of values of the label set are not. The core of tree-based algorithms is the sequential decision-making alongside the values of the respective input variables. In that sense, the data points are not considered as a “whole”, but each coordinate is independently partitioned into discrete label values. By nature, this makes it relatively ill-suited to capture inter-parameter correlations. However, this can be advantageous for high-dimensional data, where unwanted correlations (so-called co-linear behavior) often result from the limited amount of available data (as expected with AAB data). In addition, the set of successive decisions is limited. This can result in a prioritization of relevant DS-parameters, which helps to further reduce problem complexity.

Ensemble trees resample the training data—e.g., by a random draw with replacement—and train a new

decision tree on each of the draws. The resulting ensemble tree is, depending on the respective algorithm, the average of the tree ensemble (so-called bagging) but can also have a more complex algorithmic nature that includes, for example, an error-weighted average of the trees (as in boosting) [53]. Ensemble learners generally reduce the influence of noisy training data on the prediction and create more refined decision rules. However, resampling requires slightly more data which could have a negative effect for very small data sets (as is to be expected in an early experimental stage). In this work of the shelf, MATLAB functions “fitree” [54] and “fitensemble” [55] with ten surrogate splits and 10 ensemble trees were used with standard settings.

A crucial parameter of many SL methods is the uncertainty of a prediction (see section SL). More precisely, the epistemic uncertainty from the potentially erroneous assumptions of a model due to incomplete information is sought. Most ML methods do not provide an estimate of this by default because they are point estimates. However, it can be calculated as the dispersion of the prediction under slightly varying boundary conditions. To this end, varying training datasets can be created by resampling (such as jackknife bootstrapping) from the original training set as in [56]. The uncertainty then corresponds to the prediction scattering of the models trained on different samples of the training data.

Gaussian Process Regression has been introduced by Krige in the year 1951 [57] and is a probabilistic model. The core concept of GPR is to assume an underlying distribution, i.e., to treat the data as random variables. Unlike DT and TE, which learn exact values for each parameter in a function, GPR derives a probability distribution using Bayes’ rule. It updates prior knowledge (in the case of GPR, a specified distribution function) with observations (the training data) to compute a joint posterior probability distribution over all possible values. It contains information from both the prior distribution and the data set. Predictions are made through the joint distribution by weighting all possible predictions with their calculated posterior distribution. The output is the point estimate at the considered point of interest, which yields an expected value and a variance. As the latter can be naturally used as a measure of uncertainty, GPR has been a popular method for sequential learning or the closely related Bayesian optimization applications [14]. The probability

functions in GPR are commonly specified by a multivariate Gaussian distribution (in theory, other distribution can be used, too)—a so-called Gaussian process (GP)—which is defined by a mean and covariance function. The selection of GPs can incorporate a priori knowledge about boundary conditions, e.g., when periodicity, dependencies, various length scales or general trends are known. However, this is rather relevant for time series and location-dependent data and has no proximate applicability for the presented case. Furthermore, GPs control the smoothness of the (interpolated) predictions. We compared all GPs that are implemented in MATLABs statistics toolbox [58] and found that the exponential GP performed best in the SL task.

Strategies and utility functions

SL executes a strategy to select the next input by prioritizing the predictions, which are weighted by a utility function. The prioritization is conducted by—depending on whether the objective is to minimize or maximize a criterion—choosing the minimum or maximum weighted value. For simplicity, only the maximization case will be considered in the remainder of the manuscript, which can be described by Eq. (1),

$$x_{n+1} = \operatorname{argmax}(u) \quad (1)$$

where x_{n+1} is the selected next candidate and $\operatorname{argmax}(u)$ corresponds to finding the maximum utility u . Three general strategies can be distinguished. 1. Explorative strategies attempt to reduce model uncertainty by using utility functions that favor candidates with large prediction uncertainties. 2. Exploitative strategies tend to reinforce the current model perception by considering only the predicted values by the utility function (without considering uncertainties). 3. The third group is balancing between exploring and exploiting. Only 2. and 3. are greedy strategies and thus suitable for most material finding problems.

Maximum expected improvement (MEI)

The (MEI) strategy [56] purely exploits by simply selecting the next candidate according to the maximum prediction value. The utility u_i of the i -th prediction is simply:

$$u_{MEI,i} = \mu_i \tag{2}$$

where μ_i is the mean prediction of the i -th candidate.

Maximum likelihood of improvement (MLI)

The MLI strategy [56] is an explore and exploit strategy. It selects the candidate with the highest likelihood to exhibit the desired target property. In the case of normally distributed prediction, the candidate with the highest 95 percent likelihood can be determined according to Eq. (3),

$$u_{MLI,i} = Q(y_i, 0.95) = \mu_i + 1.93 * \sigma_i \tag{3}$$

where $Q(95\%)$ is the 95% quantile, μ_i is the mean prediction of the i th candidate and σ_i is the standard deviation of the i -th candidate.

MEI and MLI with maximum Euclidean distance (MEI + D and MLI + D, respectively)

At the beginning of an SL run, the predictive power of ML algorithms is relatively poor due to the small amount of training data. The data are further reduced by sampling for uncertainty estimation by DT and TE, with only a portion of the data available for each sample. This causes a situation where many candidates yield the same prediction and uncertainty value, despite the fact that their composition and processing’s are very different. Candidates that have a large average Euclidean distance to the known DS candidates differ naturally more in their design. Their choice would increase the data variability and, in turn, the predictive model’s performance will be most improved. This a-priori knowledge is naturally part of GPR, such that it outputs higher uncertainties for more distant data points. The utility function can be adjusted in a similar way by choosing the value that has the largest mean distance to the known DS candidates from a given range of prediction values. The MEI + D or MLI + D utilities were estimated according to Eq. (4) and (5), respectively.

$$u_{MEI+D,i} = \text{mean dist}(x, u_{MEI,Q(90)}) = \frac{1}{n} * \sum_{j=1}^n |x_j - x_{Q(u_{MEI},0.9),i}| \tag{4}$$

$$u_{MLI+D,i} = \text{mean dist}(x, u_{MLI,Q(90)}) = \frac{1}{n} * \sum_{j=1}^n |x_j - x_{Q(u_{MLI},0.9),i}| \tag{5}$$

where *meandist* is the mean Euclidian distance, x_j are j -th coordinates of the known training data with n samples and $x_{Q(u_{MEI},0.9),i}$ and $x_{Q(u_{MLI},0.9),i}$ are the DS coordinates of the i -th candidate with a greater than 90% quantile of the MEI or MLI utility. The MEI + D and MLI + D strategies aim at boosting the initial rounds of a SL run and hence were restricted to the 15 first iterations in the presented work. The utility was then calculated according to the MEI and MLI.

Benchmarking SL against a Random Process (RP)

Although SL is based on ML methods, classical error-based ML benchmarks typically do not apply in this context. This is because the target of SL in materials discovery is to find a candidate with—depending on the property—maximum or minimum value of a said property. In a reasonably set scenario, this goal is always achieved with zero error and is merely a matter of iterations. Although this comparison is somewhat odd from a mathematical point of view, it underscores the fact that the focus here is on the effort required to reach this threshold as a measure of performance. A common metric is the required number of experiments between (5) and (3) in Fig. 2 until a set target is reached ((6) in Fig. 2).

To determine the performance of SL methods, it is common to use simulated experiments where the ground truth labels for all data points are already known [14]. Initially, only a small fraction is provided to the SL algorithm (although more training data would be available). This is extended with one new data point from the remainder of the available data at each iteration. It is investigated which approach requires the least amount of data to achieve the goal ((6) in Fig. 2). Approaches that require less data simply lead to faster success in laboratory practice.

Thus, the goal is not to actually discover new materials using all available data, but to validate material discovery methods for scenarios where fewer labels are known (e.g., for new materials). In this approach, the generalizability is statistically demonstrated by quantifying the performance of SL methods under randomized initial conditions and then expressing it, for example, as a mean value and

standard deviation. This allows meaningful comparisons between different SL approaches. To generate randomized initial conditions in an in-progress experimental study would require significant additional effort and is unrealistic in most cases. Therefore, comparisons of performance and repeatability between different SL methods in actual material discovery are usually not possible.

This approach also differs from the classical ML approach, where generalizability is demonstrated on retained test data. However, this luxury is often not afforded in experimental science, where data are extremely limited due to costly acquisition [14].

SL is commonly compared against a Random Process (RP) (i.e., acting without a strategy and model) as a baseline benchmark. RPs consider each candidate equally likely to succeed (uniform distribution). The average number of draws necessary to find the maximum target property is 50% of the given candidates, respectively. This is the benchmark against which SL competes.

Despite the fact that this benchmark is often surpassed by SL, a significant use of SL cannot be found in practice. One reason for this may be the significantly higher effort that is caused by the sequentialization of the experimental procedure in SL. This means that from a purely functional point of view, RP can produce the desired results faster if the parallelization of experiments is more effective. In view of this situation, it is worthwhile to include further parameters for the consideration of the usefulness of SL in practice.

The specific value of the target threshold T (i.e., the property value to be exceeded) inherently affects the iteration required; the smaller the T , the fewer iterations are required for SL to succeed. From a practical perspective, relatively small deviations of the highest cement strengths contradict a special significance of a unique strength value as the target (especially considering the aleatory uncertainties of this value). To accelerate experimental progress, one can argue to reduce T , to a value that lies in the upper quantile of strengths (e.g., $T \geq f_{(c,90\%)}$) without losing much significance of the results.

Furthermore, the aspired success rate determines the number of experiments required. The relationship is simply: the higher a desired success rate, the more experiments are needed. The performance of SL at a certain success rate can be empirically determined as

the quantile of the required draws from multiple SL runs. In the laboratory practice, the required success rate is expected to be much higher than the 50% rate, which is, as mentioned above, the typical benchmark for SL.

The relationship between success rate and target threshold can be described analytically for RP as the hypergeometric cumulative distribution according to the following equation:

$$p(\text{one success} | n_{\max} \text{ draws}) = \sum_{n=1}^{n_{\max}} \frac{\binom{K}{1} \binom{N-K}{n-(k=1)}}{\binom{N}{n}} \quad (6)$$

where p corresponds to the success rate, N is the size of the population, K is the number of items with the desired characteristic in the population and n is the number of samples drawn. The threshold of success T can be defined in terms of the parameters M and x . According to Eq. (6), the success rate p has a non-linear relationship with the required draws for the case of multiple targets (where $M > 1$), i.e., the before-mentioned rule that a 50% success rate requires 50% of the possible experiments holds not for those cases. Instead, much less data are required. The exact amount further depends on the size of the population x where a greater x leads to fewer required draws n .

This means firstly that RP becomes a much tougher benchmark when T can be reduced to include more successful candidates. Secondly, a fair comparison against RP must consider the maximum available size of the population x . For example, if the DS is fragmented into several smaller DSs to parallelize SL, the RP performance must still be considered on the whole DS, as its parallelization is independent of the segmentation. In other words, benchmarks can only be compared among same size DS and the perception of the performance compared is skewed favorably to SL when smaller DSs are used.

Experiments

In this chapter, the data collection and numeric experiments are described.

Description of the data

For this study, the data related to the material discovery for acquiring samples with higher

compressive strength three data sets from five different studies were collected. This included data from [8, 59–62] about alkali-activated binders prepared from four different natural pozzolans originating from Germany and Italy, pumice stone-based natural pozzolans and granulated blast furnace slags.

First data set

The first data set included the data about four different natural pozzolans originating from Germany and Italy used to prepare geopolymers/alkali-activated binders with sodium silicate solution taken from [8, 59, 62]. These pozzolans included Rhenish trass (RT) obtained from Eifel region, Germany, Bavarian trass (BT) obtained from Nördlinger Ries, Germany, pozzolan Laziale black (PB), Ponte Lucano quarry, Tivoli, Italy and pozzolan Flegrea (PF), Campi Flegrei area, Naples, Italy. In a study by Firdous and Stephan [8, 59], these four pozzolans were subjected to react with sodium silicate solution of various silica modulus in the range of 0.4 – 1.7 ($\text{SiO}_2/\text{Na}_2\text{O}$ molar ratio) at ambient conditions (22 °C, 100% RH) to see the impact of silica modulus of alkaline solution on properties of the resultant binder. The conditions at the time of formation of pozzolan, such as the composition of magma and conditions after formation, affect the chemical and mineralogical composition of the pozzolans. Therefore, all the pozzolans had different chemical and mineralogical compositions as determined with XRF and XRD, respectively. The fineness of the material can be measured in several ways and as Blaine fineness is the quickest and most commonly used method therefore for a better comparison, the authors measured the Blaine fineness following EN 196–6 [63] and kept it in a close range ($6700 \pm 160 \text{ cm}^2/\text{g}$) for used pozzolans, whereas the d_{50} particle size measured using laser granulometry, Mastersizer 2000 of Malvern Instruments changed. In this study, sodium silicate solution in various silica moduli was used. To achieve various silica moduli, various combinations of sodium silicate solution ($\text{SiO}_2 = 30.2 \text{ wt.}\%$, $\text{Na}_2\text{O} = 14.7 \text{ wt.}\%$) and sodium hydroxide solution (3.6, 6.6, 9.2 and 11.5 mol/L) were used. Therefore, the resultant SiO_2 mol.%, Na_2O mol.%, H_2O mol.% and $\text{SiO}_2/\text{Na}_2\text{O}$ (mol/mol) are considered here for the analysis. As the change of either of them can affect the compressive strength. The alkaline solution to pozzolan ratio impacts the compressive strength;

therefore, for each pozzolan, the amount of solution required to obtained sample of good workability was determined and kept equal to 0.50, 0.75, 0.43 and 0.52 for RT, BT, PB and PF, respectively. All the paste samples prepared for compressive strength were mixed in a similar manner and cured at $22 \pm 2 \text{ }^\circ\text{C}$, 100% relative humidity (RH). Cubes with an edge length of 2 cm were used to determine compressive strength at the same loading rate for all the samples.

A further extension of the data presented above was taken from [62], where the pozzolan Laziale black and pozzolan Flegrea were additionally subjected to react with sodium and potassium hydroxide solution (9.2 and 11.5 mol/L). The prepared AAB were heat-cured at 40 °C and 100% RH for the first 28 days, followed by ambient temperature (22 °C, 100% RH) curing till 90 d. The heat-cured samples were further compared with the samples only cured at ambient conditions. To see the impact of different mineralogical compositions and fineness, the authors used a low reactive pozzolan (Bavarian trass) for further analysis [59]. In this study, the Bavarian trass was subjected to calcination and mechanical activation by burning the pozzolan at 700 °C for 3 h and by milling the pozzolan in a planetary ball mill for 5 and 10 min at a speed of 200 rpm with constant media to material ratio of 1:0.16. The calcination resulted in a reduction in the calcium carbonate content of the pozzolan and the formation of CaO, whereas the mechanical activation resulted in 23% and 43% reduction in d_{50} in comparison with un-milled Bavarian trass for 5 and 10 min milled material, respectively. For the preparation of AAB/GP samples, sodium silicate solution of silica modulus 0.707, 0.797 and 1.061 was used, whereas mixing method, rate of loading, sample size and curing conditions were kept similar as in the above study.

Second data set

The second data set is taken from [60], where the authors used pumice type natural pozzolan obtained from Taftan mountains, Iran, for the production of alkali-activated binders. Mineralogically and chemically, this pozzolan was different from those given in the first data set. The chemical composition measured following ASTM C311 [64] showed high siliceous content. The Blaine fineness of the used pozzolan was $3090 \text{ cm}^2/\text{g}$ which is lower than the fineness of pozzolans used in the first data set. Sodium silicate

solutions of different silica modulus in the range of 0.3 – 0.9 were used, and $\text{Na}_2\text{O}/\text{Al}_2\text{O}_3$ molar ratio of the system was also changed in the range of 0.77 – 1.23. To achieve the desired silica modulus, a combination of sodium hydroxide and sodium silicate solution was used. For determination of compressive strength, 2 cm^3 samples were cured at $25\text{ }^\circ\text{C}$ and 95% RH.

Third data set

The third data set was published by Tänzler [61]. In this work, artificial glasses were prepared by blending and remelting mixes of one GGBFS and several oxides. The single melting experiments were performed between $1550\text{ }^\circ\text{C}$ and $1650\text{ }^\circ\text{C}$ in a nitrogen atmosphere, and the material was granulated at 3 bar in water of $10\text{ }^\circ\text{C}$. The granulate was ground on a laboratory ball mill. The laser granulometry with a Horiba LA-300 revealed values for d_{50} from 8.5 to $16\text{ }\mu\text{m}$, and the Blaine fineness according to DIN 66,126 [65] and EN 196–6 [66] was found to be 4120–4800 cm^2/g . The chemical constituents of the synthesized slags were determined following EN ISO 11,885 [67] based on ICP-OES in the case of Al_2O_3 , CaO, Fe, K_2O , MgO, MnO, Na_2O , S and TiO_2 . Furthermore, SiO_2 was determined gravimetric according to EN 196–2 [68]. Due to the non-oxidizing conditions while glass manufacturing, the oxidation state of Fe and S is unknown. For the current evaluation, it was assumed that both elements are fully oxidized.

Sodium hydroxide (2 mol/kg), sodium waterglass ($M_s = 1$) and two potassium waterglass ($M_s = 1$ and 2) solutions were used as activators. The solution–powder ratio was between 0.374 and 0.432. Pastes were prepared by a hand mixer in 30 s. The paste-filled molds for 2-cm cubic samples were treated on a vibration table and thereafter covered with foil until the demolding after $23 \pm 1\text{ h}$. The demolded cubes were stored at $20\text{ }^\circ\text{C}$ and 100% RH.

Summary of design space—features and target property

Based on these three data sets with a total of 131 samples, the DS was constructed. The 22 features used in this study included for the powdered precursor the 13 parameters of the chemical composition in wt.% (SiO_2 , Al_2O_3 , Fe_2O_3 , MnO, MgO, CaO, Na_2O , K_2O , TiO_2 , P_2O_5 , SO_3 , Cl and loss on ignition). For

characterizing the granulometry, three features were used, namely the Blaine fineness [cm^2/g] as well as d_{90} and d_{50} particle size [μm]. The composition of the alkaline solution is given with four features in terms of SiO_2 mol.%, Na_2O or K_2O mol.% and H_2O mol.%. Furthermore, the curing temperature and the solution–powder ratio were used. All features were normalized so that their respective mean was zero and the standard deviation was scaled to one.

The resultant compressive strength [MPa] for 2.2 cm^3 samples at the age of 28 days is the sought-after target property.

Investigated scenarios

One way to visualize the SL task is to represent the DS in T-SNE coordinates. T-SNE [69] is a dimension reduction algorithm much like the well-known principal component analysis (PCA). These allow to represent higher dimensional vector spaces (like our DS) in a lower dimensional form (e.g., in two dimensions). The dimension reduced space has no physical meaning anymore, but allows to analyze the characteristic distribution of the data points, i.e., points that are close to each other have similar feature values. Trends in the distribution of the data, such as the relationship between feature values and materials properties can be inferred (with some uncertainty). In contrast with PCA, T-SNE employs nonlinear dimensionality reduction. This has the advantage of preserving the relationships between neighboring data points, while reducing the large interdimensional distances that occur especially in sparsely populated high-dimensional spaces.

Figure 3 shows the distribution of our data set in T-SNE coordinates: on the left, subsets 1–3 from Sect. 3.1.1 to 3.1.3 are represented by different colors; on the right, the distribution of compressive strength is shown as a color scale.

The clear distinction of the data on the left side indicates that the difference in material composition is well represented by the features. The global optimum of the strength distribution is found in the third data set (cf. pink dot in Fig. 3, right). The strength distribution within the three materials appears uniform with relatively clear trends toward the respective local maxima, suggesting a good relationship between features and material properties. We compared SLs ability to find the higher-strength materials in various scenarios (see Table 2).

Figure 3 Input data set in TSNE coordinates; left: Three clusters from K-Means clustering ($k = 3$) with each cluster corresponding to one data set (cf. Section 3.1.1–3.1.3); right: Colored by compressive strength in MPa.

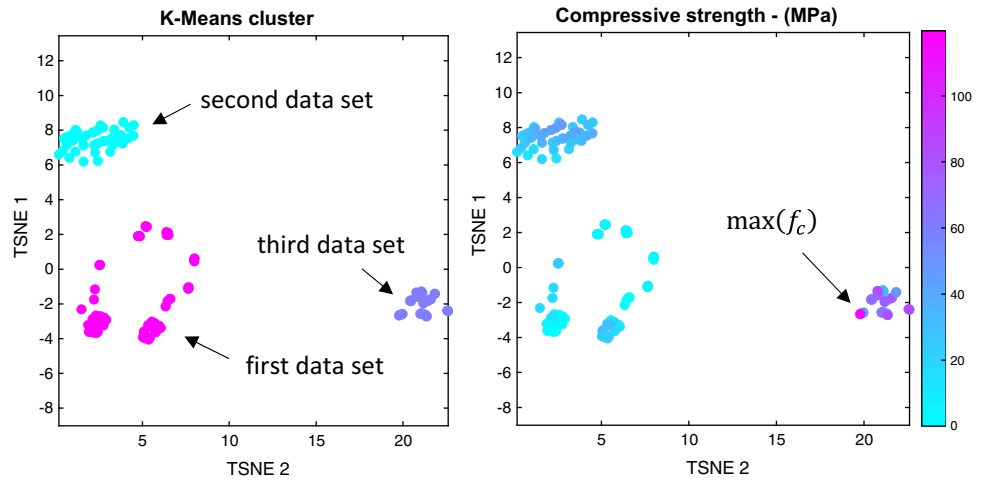


Table 2 Influencing factor on the success of SL and its variations that are investigated in this paper

Influencing factor	Investigated variation
1 Prediction algorithm and utility function	GPR, DT, TE MEI, MLI, MEI + D, MLI + D
2 DS complexity	Segmented DS vs. joint DS
3 Initial training set size	4, 8, 12
4 Target threshold T	$T = \text{maximum strength}$ vs. $T \geq 90\%$ quantile strength
5 Success rate	50% vs. 97%, i.e., performance in 15/30 and 29/30 SL runs, respectively

First, combinations of the three prediction algorithms with four utility functions were investigated.

Second, the complexity of the SL task was altered, with SL performed in a common DS containing the 131 data points and a segmented DS. The segmentation breaks the complex optimization problem into separate, expectedly simpler problems—potentially creating better predictability and possibly further speed-up. The segmentation was achieved with k-means clustering (with $k = 3$) of the data in two-dimensional T-SNE coordinates using the standard functions in MATLAB, respectively [70, 71]. This created clusters that correspond to the three data resources (comparison Fig. 3, left). In addition, the segmentation of the problems allows parallelization, since the SL runs could be executed in batches consisting of the segments.

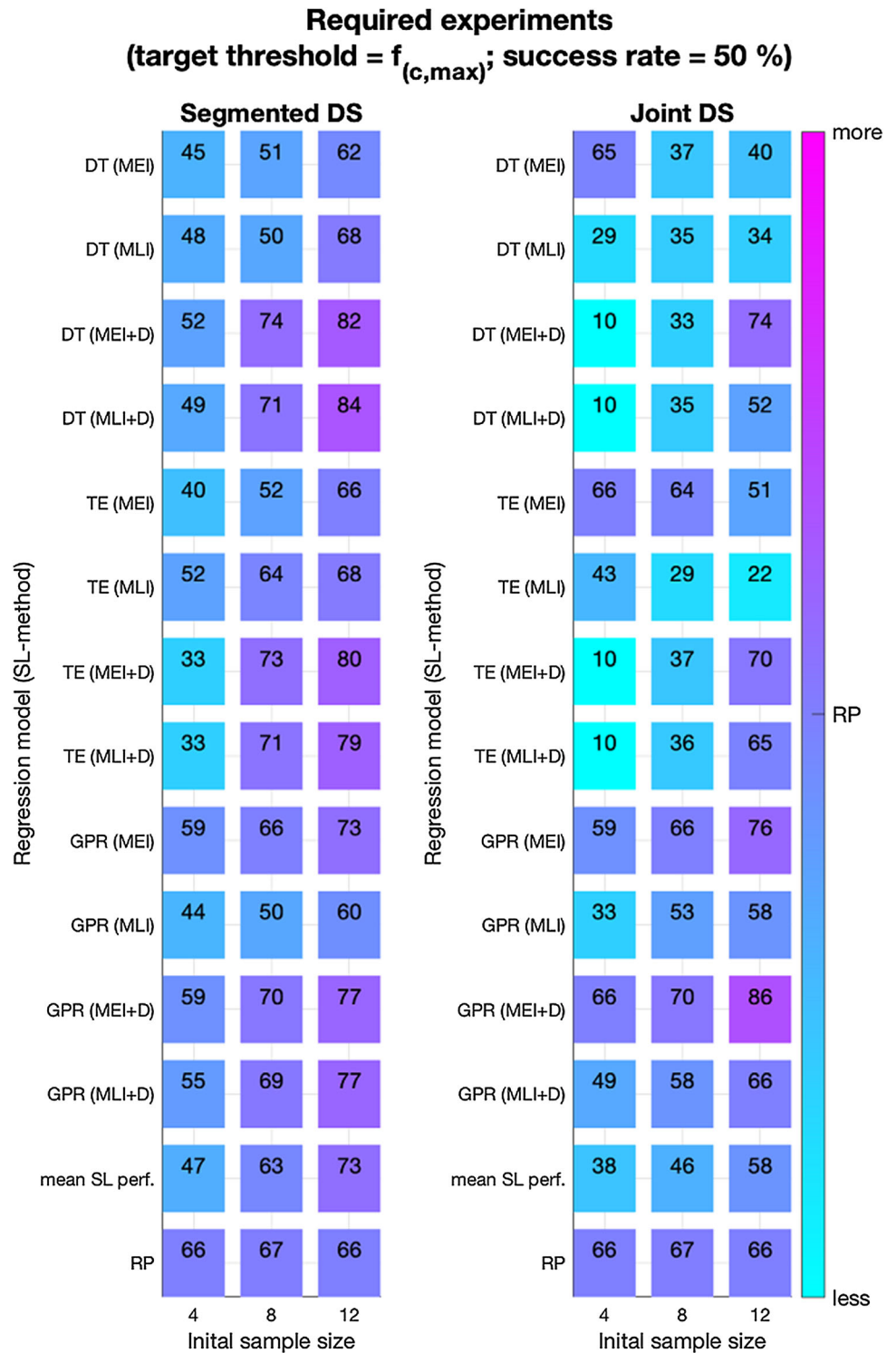
Third, the initial training set size was varied between four and twelve. The minimal initial training set size of four was restricted by two sub-samplings in the TE with uncertainties algorithm (each leaving out one candidate) and a minimum of two points in a sensible regression problem. The upper boundary of

the initial training set was restricted to 12 to allow 30 different samples to be drawn from the lower half (as per restriction the 50% quantile restriction discussed below) of the smallest data set segment.

Fourth, the task of the SL has been varied in terms of the success threshold between finding the absolute maximum within one DS and finding one of the above 90% quantile compressive strengths within one DS.

Fifth, the benchmarking was carried out statistically in terms of a 50% (Figures 4 and 6) and 97% (Figures 5 and 7) success rate. The former is the usual benchmark indicating the average performance and the latter indicates the robustness, which has special importance in scientific studies. The success rates were estimated from 30 SL runs with randomly sampled initial training sets. These initial samples were restricted to the lower half (50% quantile) of the strengths to enforce a significant improvement in the discovery process. Since extrapolation is much more difficult than interpolation for many learning algorithms, it is hoped that this will provide a more realistic understanding of SL performance. It is

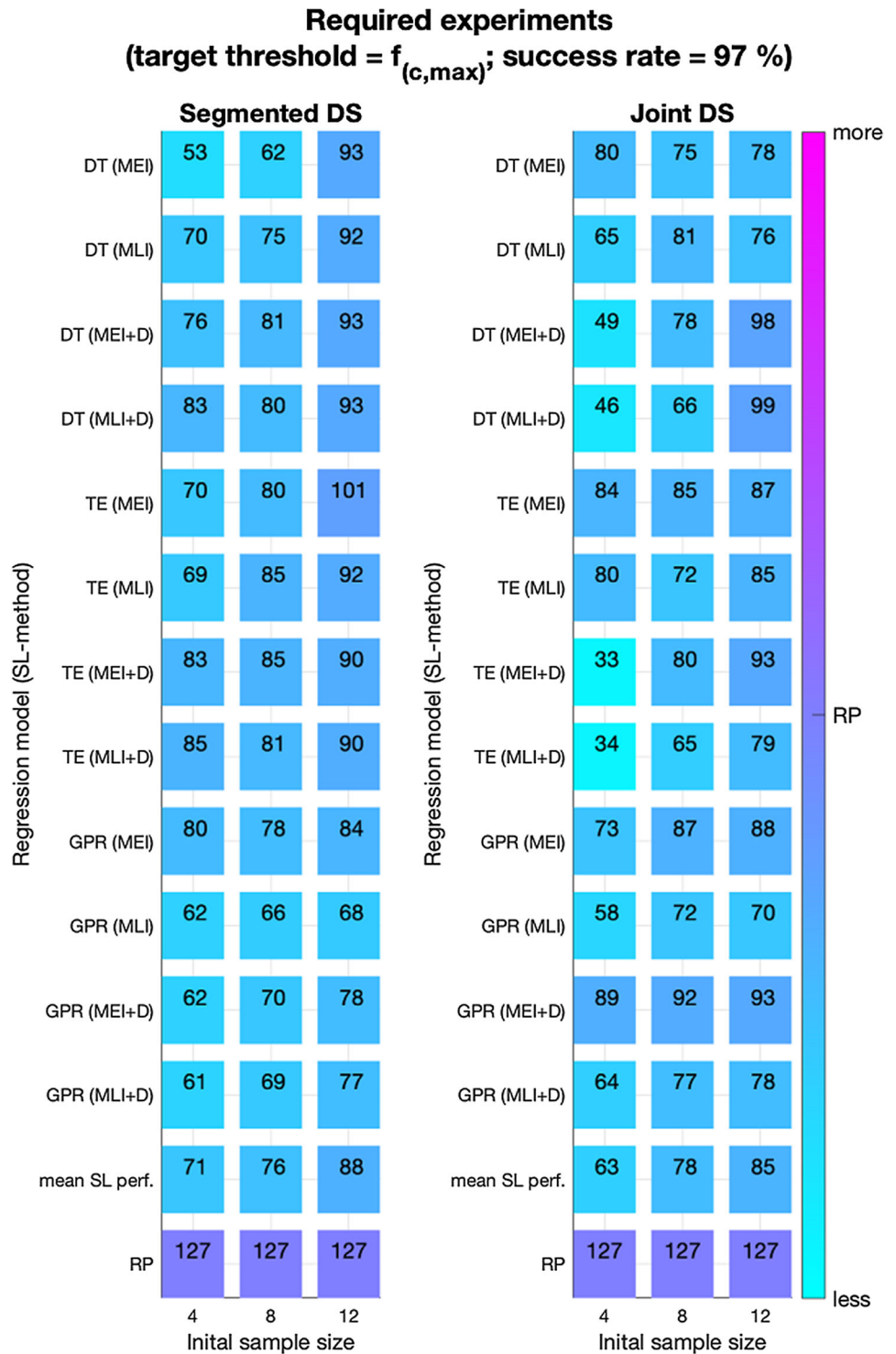
Figure 4 Performance of 12 SL algorithms vs. RP in terms of required experiments for finding the maximum strength with a 50% success rate (left for segmented DS, right for joint DS).



acknowledged that this restriction has a considerable influence on the performance of SL. However, in practice, whether the increased performance would be achieved remains largely unknown and depends

on the given candidate group (see further details in [47]).

Figure 5 Performance of 12 SL algorithms vs. RP in terms of required experiments for finding the maximum strength with a 97% success rate (left for segmented DS, right for joint DS).

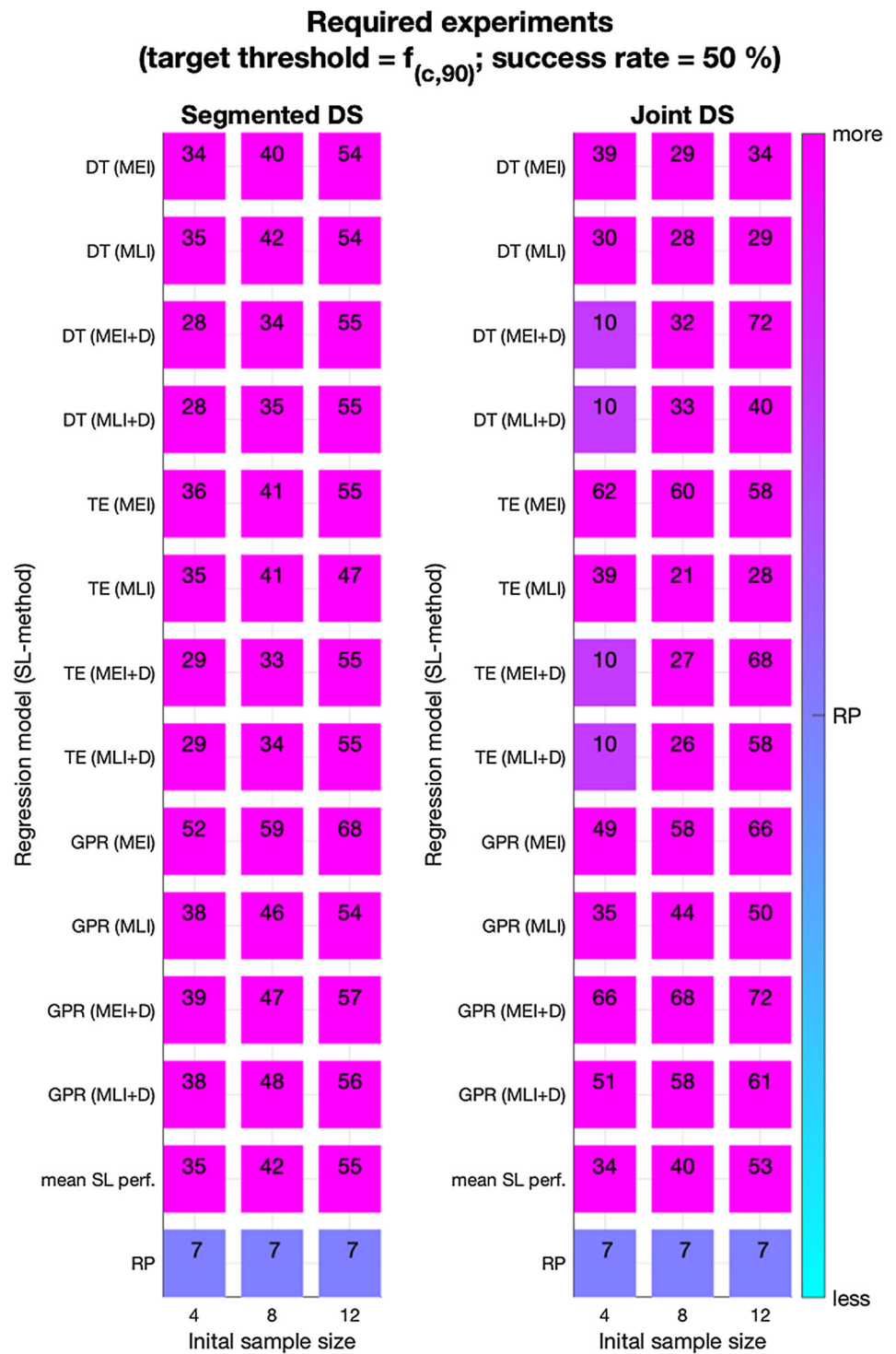


Results and discussion

The investigated scenarios include the twelve SL algorithms, which are used in two different DS, initialized with three different initial training set sizes to achieve two different target thresholds, respectively,

which are statistically analyzed in terms of a 50% and 97% success rate. This results in a total of 288 SL results which are shown in Figs. 4, 5, 6 and 7. The second to last line in each figure shows the average performance of the SL algorithms, and the last line

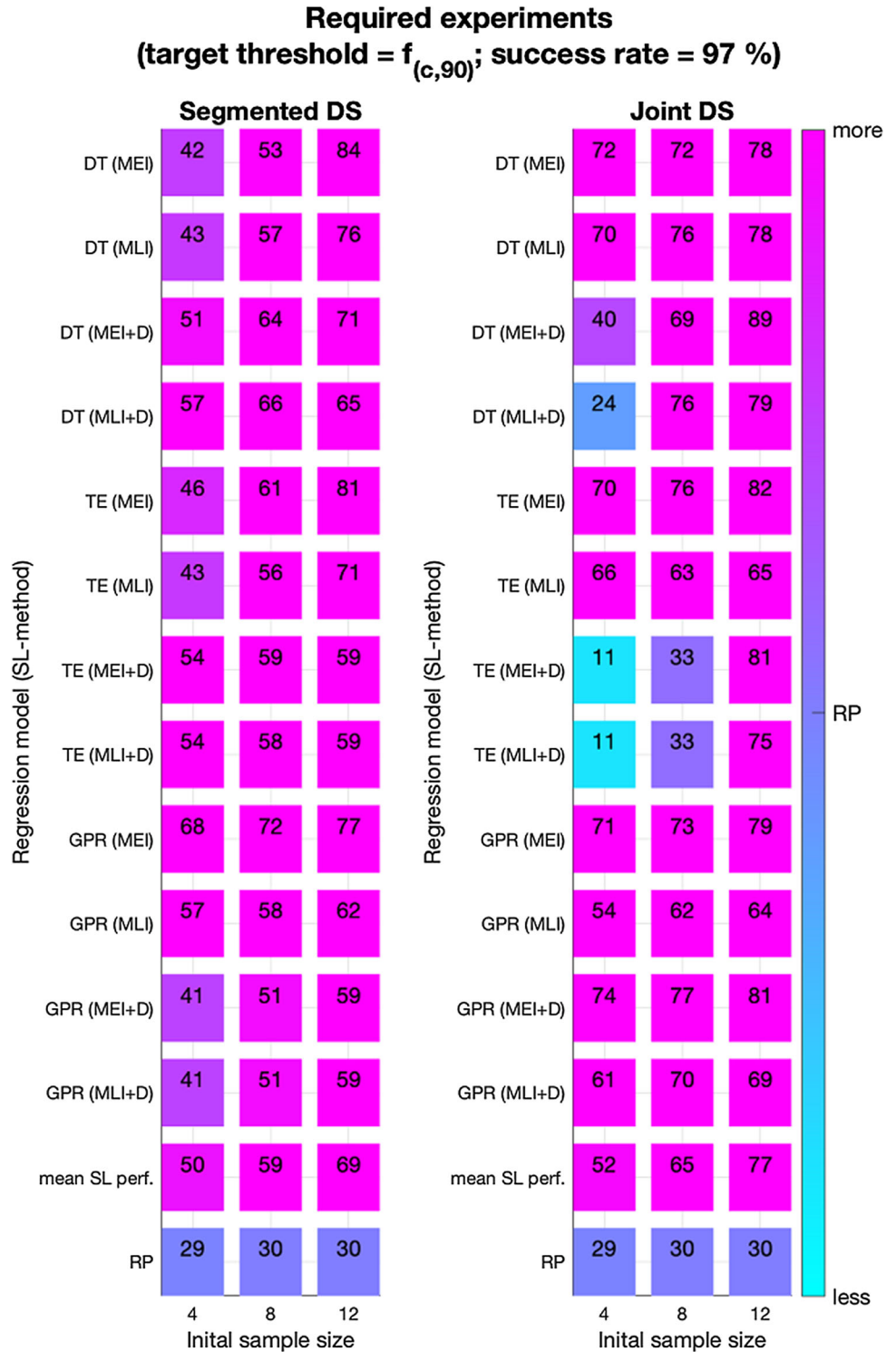
Figure 6 Performance of 12 SL algorithms vs. RP in terms of required experiments for finding one candidate above the 90 percent compressive strength quantile with a success rate of 50% (left for segmented DS, right for joint DS).



shows the RP performance. The results are shown as the number of required experiments for the specific success rate (cf. Section 2.2.3). Higher numbers mean longer SL runs and hence worse performance. The results on the left correspond to the cumulative results from the three segmented DS. The results on

the right correspond to the SL performance in the joint DS. Both were benchmarked against the RP in the joint DS (as mentioned at the end of Sect. 2.2.3). To allow a quick comparison between different scenarios, the results were color coded, with pink-colored boxes indicating worse performance and

Figure 7 Performance of 12 SL algorithms vs. RP in terms of required experiments for finding one candidate above the 90 percent compressive strength quantile with a success rate of 97% (left for segmented DS, right for joint DS).



turquoise-colored boxes indicating better performance than that of the respective RP. Different target thresholds are compared with Figs. 4 and 5 for finding the maximum strength in the respective DS and

Figs. 6 and 7 for finding strengths above the 90% quantile.

When considering the average performance of the SL algorithms in Figs. 4, 5, 6 and 7, it can be found that SL in segmented DS and joint DS is in a

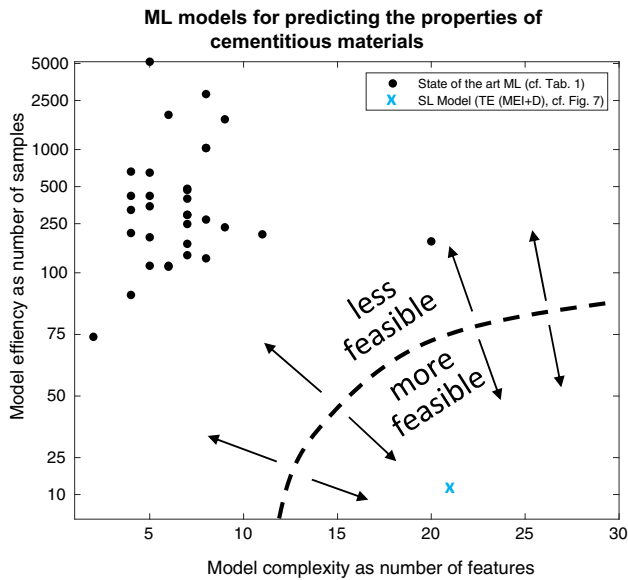


Figure 8 A comparison between state-of-the-art ML models and SL for predicting cementitious material properties in terms of feasibility in laboratory practice as a function of model complexity and model efficiency.

comparable range of numbers. This has practical relevance because—as mentioned in introduction—the former can be parallelized and thus be quicker in practice. Generally, it seems to pay to start SL as early as possible with few initial training data. The higher cost of collecting a larger data set in the initial phase is sometimes complemented by even longer run times in terms of more experiments needed in the SL run. This demonstrates how important it can be for the success of SL to explore the DS quickly and that a larger set of weak initial assumptions can cause a confirmation bias that negatively affects the course of SL.

The higher performance of SL over RP is more likely at higher success rates (see Figs. 5 and 7), which speaks for the higher robustness of SL. This is practically relevant because it shows that, depending on the success rate requirements, approaches with higher robustness perform better than approaches with the more commonly compared average performance. When comparing RP, it seems that in Figs. 4 and 5, RP is a benign benchmark with relatively weak performance and low robustness and that the lower target threshold in Figs. 6 and 7 has a strong favorable impact on performance and robustness—making RP an ever more challenging benchmark.

In contrast, lowering the target threshold improves the average SL performance only slightly. Since SL

requires training data, it must have a performance offset (in terms of the minimum required experiments) to its disadvantage at the beginning—the more initial training data, the greater. RP, however, can just be lucky with the first draw. SL's only chance of compensating for this disadvantage is to succeed quickly and repeatedly. This in itself is a remarkably difficult task since the higher-strength candidates are in different regions of the high-dimensional DS than the initial training data (cf. Figure 3, right); this means that SL has to extrapolate considerably. The MEI + D and MLI + D utility functions accelerate SL's DS exploration to the extent that makes it possible to outperform RP even for a lower target threshold. This is observed in Fig. 7 for the joint DS in the case of four initial training sets for DT and MLI + D as well as TE and MEI + D and MLI + D. Adding the distance to the utility function essentially extends the exploratory component of the algorithms. However, this does not seem to work for GPR. This can be explained by its tendency to revert predictions to the function mean for more distant points. This means that candidates that have a greater distance (which is exploited by MEI + D and MLI + D) will get lower prediction values. However, since only the 10% highest predictions are considered, candidates with smaller distances are more likely to be chosen here, resulting in less exploration. The exception is the segmented DS, where the distances between the points are smaller. Here, MEI + D and MLI + D improve the robustness of GPR in the low target threshold scenario (cf. Figure 7). The performance of GPR is clearly best when combined with the MLI strategy. DT and TE, on the other hand, benefit from MLI over MEI, especially in the larger joint DS. Looking at the results, the uncertainty-based and exploratory approaches tend to be more successful in larger DS—as would be expected. Adding more training data also improves DT and TE performance in some cases. However, in the lower success threshold scenario, this is not enough to match RP. In addition, MEI + D and MLI + D perform poorly when more initial training data are added. The overall best approach is consistently TE with MEI + D or MLI + D with four as the initial training set size in the joint DS.

Figure 8 illustrates how the relationship between model complexity and model efficiency (note the non-linear scaling of the Y-axis) affects feasibility for laboratory practice. Here the characterization of

AABs, on the one hand, involves considering many parameters that potentially vary widely between candidate materials (as discussed in the literature review Sect. 2.1). On the other hand, collecting data is time-consuming and usually must be repeated for each new material. Both aspects limit the applicability of current ML models in practice. Although the specific course of the feasibility boundary depends on the respective laboratory capacity and the complexity of the underlying material, it is clear that the presented SL approach (blue cross) is much more effective than the state-of-the-art ML models (black dots).

Conclusions

Considering the pressing climate crisis, the large CO₂ footprint of OPCs and the complexity of new binders, new materials discovery methods are urgently needed. The amount of available data is one of the major bottlenecks in using ML for this. Existing approaches require large amounts of data and generalize potentially poorly due to their coarse input parameters. We argue for including much more detailed information in ML models and showed how useful material models can still be generated with very few data sets (just four to get started). To this end, we apply an SL approach to an AAB dataset. We analyze some of the issues that have arisen for use in laboratory practice. The results show that the discovery of new precursors for construction binders can be accelerated using SL. Specifically, the acceleration is achieved in terms of fewer experiments required to predict materials with the desired properties. Figure 8 shows the tremendous improvement compared to state-of-the-art ML models from the literature. To achieve this result, SL's purely statistically motivated benchmarking approaches were adapted to aspects of laboratory practice, e.g., by lowering the target threshold and comparing results at higher success rates. From this, the following conclusions can be drawn for SL in laboratory practice. Despite very complex data, it seems promising to integrate SL as early as possible in the experimental program. It has been shown that ML-based SL has the potential to outperform classical GPR. The parallelization of SL runs in three batches had a relatively small average influence on the performance in the example presented. Regarding high-throughput experiments, such parallelization procedures are vital research topics. Lowering the target

threshold led to a surprisingly small boost of most SL procedures and a significant acceleration of the RP. Even if these considerations are of a rather intellectual nature, because in practice, the actual distribution of the properties is unknown—it still shows the advantages that RP can have. Incorporating randomized effects—i.e., effects that have nothing to do with the deterministic understanding of the respective ML prediction—could be a key to further acceleration for SL. This has been approached with some success in this publication by adding a distance in the DS to the utility function. TE with MEI + D and MLI + D achieved an outstanding performance with only eleven required experiments to find higher-strength cements with a 97% success rate (cf. Figure 7, right). In fact, this is nearly 60 times less training data than the models in Table 1 require on average, while more than three times as many material composition features were used. This approach seems to break with the paradigm in data science that more training data produces better models. This is quite significant because it lowers the threshold of adoption for laboratory practice.

The work presented in this paper leads us to the conclusion that further research in this field is needed as there are still many open questions—e.g., on the influence of the DS complexity (in terms of very large DSs) or the exact interaction between the amount of initial training data and the success of the approach. It would also be interesting to undertake a systematic study of the added hyper-parameters of the utility functions with distance.

In addition, the relevance of input features for SL is an important question for future research. On the one hand, adding more features can help to better characterize the material, which can lead to even better predictability and faster development cycles. The same is true for omitting superfluous data. On the other hand, some information is currently unusable because it is either not uniformly available or features are simply not easily extracted (e.g., X-ray diffraction data). In turn, the use of such information increases the complexity of data analysis and acquisition, which can lead to longer development cycles. Furthermore, a deeper understanding of the crucial parameters would help to check SL for plausibility. After all, confidence in the method is one of the main prerequisites for its application in practice. To answer the question of how SL methods are to be configured in a wide range of tasks in a variety of scenarios,

meta-learning (e.g., as proposed in [72]) could be considered. To make the required highly complex data available, a powerful (possibly semantic) data management is needed that goes far beyond the already extensive data collection by Xie et al. [46].

While the objective of this study is to find alkali-activated binders with high compressive strength, the method can potentially be applied to other properties of the building materials, such as durability or workability. The potential of the investigated methods goes beyond material discovery and could soon enable cost-effective and ecological re-engineering of existing binders or even “materials by design” applications in construction.

Acknowledgements

We would like to thank Philipp Benner from the Federal Institute for Materials Research and Testing for the many valuable discussions on the background of the mathematical methods and Julia Ling from Citrine Informatics for inspiring this work through her publications and online lectures. The authors are thankful to the German Academic Exchange Service and Higher Education Commission of Pakistan (Programme ID: 57245990) for funding Rafia Firdous.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Data availability

The data and source code necessary to reproduce the results are available under the link: <https://github.com/BAMcvoelker/Sequential-Learning-for-AAB-discovery>.

Declarations

Conflict of 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.

Supplementary Information: The online version contains supplementary material available at <http://doi.org/10.1007/s10853-021-06324-z>.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Supplementary Information: The online version contains supplementary material available at <http://doi.org/10.1007/s10853-021-06324-z>.

References

- [1] Hatfield, Ashley (2021) “National Minerals Information Center,” [Online]. Available: <https://www.usgs.gov/centers/nmic/cement-statistics-and-information>. [Accessed 20 04 2021].
- [2] Andrew RM (2018) “Global CO₂ emissions from cement production,” *Earth System Science Data*, p. 195–217.
- [3] Dell R (2021) “Making the Concrete and Steel We Need Doesn’t Have to Bake the Planet,” [Online]. Available: <https://www.nytimes.com/2021/03/04/opinion/climate-change-infrastructure.html>. [Accessed 29. March 2021].
- [4] World Business Council for Sustainable Development, Cement Sustainability Initiative. , “Technology Roadmap - Low-Carbon Transition in the Cement Industry,” International Energy Agency, France, 2018.
- [5] Provis JL, van Deventer J (2014) *Alkali Activated Materials, State-of-the-Art Report, RILEM TC 224-AAM*. Springer, Dordrecht, Heidelberg, New York, London
- [6] Shobeiri V, Bennett B, Xie T and Visintin P (2021) “A comprehensive assessment of the global warming potential of geopolymers concrete,” *J Cleaner Production*, 297, p. <https://doi.org/10.1016/j.jclepro.2021.126669>.

- [7] Firdous R, Stephan D, Djobo JNY (2018) Natural pozzolan based geopolymers: A review on mechanical, microstructural and durability characteristics. *Constr Build Mater* 190:1251–1263. <https://doi.org/10.1016/j.conbuildmat.2018.09.191>, 27 September
- [8] Firdous R and Stephan D (2019) “Effect of silica modulus on the geopolymerization activity of natural pozzolans,” *Construction and Building Materials*, pp. 31–43, 219; <https://doi.org/10.1016/j.conbuildmat.2019.05.161>.
- [9] Damtoft JS, Lukasik J, Herfort D, Sorrentino D, Gartner EM (2008) Sustainable development and climate change initiatives. *Cem Concr Res* 38(2):115–127. <https://doi.org/10.1016/j.cemconres.2007.09.008>
- [10] Guess M (2019) “It’ll cost \$45 billion, but Germany proposes to eliminate coal in 19 years. *Ars Technica*,” [Online]. Available: <https://arstechnica.com/tech-policy/2019/01/itll-cost-45-billion-but-germany-proposes-to-eliminate-coal-in-19-years/?comments=1&pos>. [Accessed 27 03 2021].
- [11] Mathieson JG, Somerville MA, Deev A and Jahanshahi (2015) “Utilization of biomass as an alternative fuel in ironmaking.,” in *Iron Ore*, vol. 42, Elsevier, p. 581–613.
- [12] Manz OE (1997) Worldwide production of coal ash and utilization in concrete and other products. *Fuel* 76(8):691–696
- [13] Provis JL (2018) Alkali-activated materials. *Cem Concr Res* 114:40–48. <https://doi.org/10.1016/j.cemconres.2017.02.009>
- [14] Lookman T, Balachandran PV, Xue D, Yuan R (2019) Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *NPJ Comput Mater* 21(5):2–17
- [15] Reyes K and Powell WB (2020) “Optimal Learning for Sequential Decisions in Laboratory Experimentation,” *SIAM Review* (preprint), pp. 1–47.
- [16] Addy SA, Koshla NP (1986) Computer Aided Design of Portland Cement Concrete Mixes. *Transp Res Rec* 1062:63–69
- [17] Naderpour H, Rafiean AH and Fakharian P (2018) “Compressive strength prediction of environmentally friendly concrete using artificial neural networks,” *J Build Eng*, pp. 213–219; <https://doi.org/10.1016/j.jobe.2018.01.007>.
- [18] Chaabene WB, Flah M and Nehdi ML (2020) “Machine learning prediction of mechanical properties of concrete: Critical review,” *Construction and Building Materials*, vol. 260, p. <https://doi.org/10.1016/j.conbuildmat.2020.119889>.
- [19] Thomas RJ and Peethamparan S (2017) “Stepwise regression modeling for compressive strength of alkali-activated concrete,” *Construction and Building Materials*, p. 141 (2017) 315–324; <https://doi.org/10.1016/j.conbuildmat.2017.03.006>.
- [20] Bagheri A, Nazari A, Sanjayan J (2019) “The use of machine learning in boron-based geopolymers: Function approximation of compressive strength by ANN and GP,” *Measurement*, p. Measurement 141(2019):241–249. <https://doi.org/10.1016/j.measurement.2019.03.001>, 18 April
- [21] Gomaa E, Han T, ElGawady M, Huang J and Kumar A (2020) “Machine learning to predict properties of fresh and hardened alkali-activated concrete,” *Cement and Concrete Composites*, p. 115 (2021) 103863; <https://doi.org/10.1016/j.cemconcomp.2020.103863>,
- [22] Ahmad A, Farooq F, Niewiadomski P, Ostrowski K, Aslam AAF and Alyousef R (2021) “Prediction of Compressive Strength of Fly Ash Based Concrete Using Individual and Ensemble Algorithm,” *Materials (Basel)*. 14(4):794. <https://doi.org/10.3390/ma14040794>
- [23] Van Dao D, Ly H, Trinh S, Le T and Pham B (2019) “Artificial intelligence approaches for prediction of compressive strength of geopolymer concrete,” *Materials (Basel)*, vol. 12, p. <https://doi.org/10.3390/ma12060983>.
- [24] Kandiri A, Sartipi F, Kioumars M (2021) Predicting compressive strength of concrete containing recycled aggregate using modified ANN with different optimization algorithms. *Appl Sci* 11(485):1–19. <https://doi.org/10.3390/ap11020485>, 6 January
- [25] Deng F, He Y, Zhou S, Yu Y, Cheng H, Wu X (2018) Compressive strength prediction of recycled concrete based on deep learning. *Constr Build Mater* 175:562–569. <https://doi.org/10.1016/j.conbuildmat.2018.04.169>
- [26] Xu J, Zhao X, Yu Y, Xie T, Yang G and Xue J (2019) “arametric sensitivity analysis and modelling of mechanical properties of normal- and high-strength recycled aggregate concrete using grey theory, multiple nonlinear regression and artificial neural networks,” *Construction and Building Materials*, vol. 211, pp. 479–491, <https://doi.org/10.1016/j.conbuildmat.2019.03.234>.
- [27] Mohammadi E, Behnood A (2018) Application of soft computing methods for predicting the elastic modulus of recycled aggregate concrete Input layer Output layer. *J Clean Prod* 176:1163–1176. <https://doi.org/10.1016/j.jclepro.2017.11.186>
- [28] Behnood A and Golafshani E (2020) “Machine learning study of the mechanical properties of concretes containing waste foundry sand,” *Constr. Build. Mater.* vol. 243, p. <https://doi.org/10.1016/j.conbuildmat.2020.118152> 118152.
- [29] Gholampour A, Mansouri I, Kisi O (2020) Evaluation of mechanical properties of concretes containing coarse recycled concrete aggregates using multivariate adaptive regression splines (MARS), M5 model tree (M5Tree) and least squares support vector regression (LSSVR) models.

- Neural Comput & Applic 32:295–308. <https://doi.org/10.1007/s00521-018-3630-y>
- [30] Yu Y, Li W, Li J and Nguyen TN (2018) “A novel optimised self-learning method for compressive strength prediction of high performance concrete,” *Construction and Building Materials*, pp. 229–247; <https://doi.org/10.1016/j.conbuildmat.2018.06.219>,
- [31] Douma OB, Boukhatem B, Ghrici M and Tagnit-Hamou A (2017) “Prediction of properties of self-compacting concrete containing fly ash using artificial neural network,” *Neural Comput. Appl.* vol. 28, pp. 707–718.
- [32] Kao CY, Shen CH, Jan JC and Hung SL (2018) “A Computer-Aided Approach to Pozzolanic Concrete Mix Design,” *Advances in Civil Engineering*, pp. 1–15; <https://doi.org/10.1155/2018/4398017>
- [33] Ziolkowski P, Niedostatkiewicz M (2019) Machine learning techniques in concrete mix design. *Mater* 12(1256):1–16. <https://doi.org/10.3390/ma12081256>, 17 April
- [34] Golafshani EM, Behnood A and Arashpour M (2020) “Predicting the compressive strength of normal and High-Performance Concretes using ANN and ANFIS hybridized with Grey Wolf Optimizer,” *Constr Build Mater.* pp. 1–14
- [35] Behnood A, Behnood V, Modiri M, Esat K (2017) Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm. *Constr Build Mater* 142:199–207. <https://doi.org/10.1016/j.conbuildmat.2017.03.061>
- [36] Han Q, Gui JXC, Lacidogna G (2019) A generalized method to predict the compressive strength of high-performance concrete by improved random forest algorithm. *Constr Build Mater* 226:734–742. <https://doi.org/10.1016/j.conbuildmat.2019.07.315>
- [37] Feng DC, Liu ZT, Wang XD, Chen Y, Chang JQ, Wei DF and Jiang ZM (2020) “Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach,” *Construction and Building Materials*, pp. 1–11; <https://doi.org/10.1016/j.conbuildmat.2019.117000>
- [38] Hameed M, AlOmar M, Baniya W and AlSaadi MA (2021) “Prediction of high-strength concrete: high-order response surface methodology modeling approach,” *Engineering with Computers*, pp. <https://doi.org/10.1007/s00366-021-01284-z>
- [39] Lapeyre J, Han T, Wiles B, Ma H, Huang J, Sant G and Kumar A (2021) “Machine learning enables prompt prediction of hydration kinetics of multicomponent cementitious systems,” *scientific reports*, pp. 1–16; <https://doi.org/10.1038/s41598-021-83582-6>
- [40] Moro C, Fil HE, Francioso V and Velay-Lizancos M (2020) “Influence of water-to-binder ratio on the optimum percentage of nano- TiO₂ addition in terms of compressive strength of mortars: A laboratory and virtual experimental study based on ANN model,” *Constr Build Mater*, pp. 1–11; <https://doi.org/10.1016/j.conbuildmat.2020.120960>
- [41] Zhang J, Ma G, Huang Y, sun J, Aslani F and Nener B (2019) “Modelling uniaxial compressive strength of lightweight self-compacting concrete using random forest regression,” *Constr Build Mater*, pp. 713–719; <https://doi.org/10.1016/j.conbuildmat.2019.03.189> .
- [42] Mundher Z, Deo R, Hilal A, Abd A, Cornejo L, Salcedo-sanz S, Nehdi M (2018) Software Predicting compressive strength of lightweight foamed concrete using extreme learning machine model. *Adv Eng Softw* 115:112–125. <https://doi.org/10.1016/j.advengsoft.2017.09.004>
- [43] Asteris G, Kolovos K (2019) Self-compacting concrete strength prediction using surrogate models. *Neural Comput Appl* 31:409–424. <https://doi.org/10.1007/s00521-017-3007-7>
- [44] Gupta T, Patel K, Siddique S, Sharma R and Chaudhary S (2019) “Prediction of mechanical properties of rubberised concrete exposed to elevated temperature using ANN,” *Measurement*, vol. 147.
- [45] Bachir R, Mohammed A, Habib T (2018) Using artificial neural networks approach to estimate compressive strength for rubberized concrete. *Period Polytech Civ Eng* 62:858–865. <https://doi.org/10.3311/PPci.11928>
- [46] Xie T, Visintin P, Zhao X and Gravina R (2020) “Mix design and mechanical properties of geopolymer and alkali activated concrete: Review of the state-of-the-art and the development of a new unified approach,” *Constr Build Mater*, p. 256 (2020) 119380; <https://doi.org/10.1016/j.conbuildmat.2020.119380>
- [47] Kim Y, Kim E, Antono E, Meredig B and Ling J (2020) “Machine-learned metrics for predicting the likelihood of success in materials discovery,” *npj Computational Materials*, vol. 6, no. 131
- [48] Bondar D, Lynsdale CJ, Milestone NB, Hassani N and Ramezani-pour AA (2011) “Effect of type, form and dosage of activators on strength of alkali-activated natural pozzolans,” *Cem. Concr. Compos.* 33 (2)
- [49] Shi C, Krivenko PV and Roy D (2006) *Alkali-Activated Cements and Concretes*, Abingdon: Taylor & Francis Group.
- [50] Firdous R, Stephan D (2020) Impact of the mineralogical composition of natural pozzolan on properties of resultant geopolymers. *J Sustainable Cem-Based Mater* 176(1):1–16. <https://doi.org/10.1080/21650373.2020.1809028>
- [51] Lindley DV (1956) On a measure of the information provided by an experiment. *Ann Math Stat* 27:986–1005. <https://doi.org/10.1214/aoms/1177728069>
- [52] Quinlan RJ (1986) *Induction of Decision Trees*, Boston. Kluwer Academic Publishers, USA

- [53] Opitz D and Maclin D (1999) “Popular Ensemble Methods: An Empirical Study,” *Journal of Artificial Intelligence Research*, p. <https://doi.org/10.1613/jair.614>
- [54] Mathworks (2021) “fitrtree - Fit binary decision tree for regression,” [Online]. Available: <https://www.mathworks.com/help/stats/fitrtree.html>. [Accessed 29 03 2021].
- [55] Mathworks (2021) “fitrensemble - Fit ensemble of learners for regression,” [Online]. Available: https://www.mathworks.com/help/stats/fitrensemble.html?searchHighlight=fitrensemble&s_tid=srchtitle. [Accessed 29 March 2021].
- [56] Ling J, Hutchinson M, Antono E, Paradiso S, Meredig B (2017) High-dimensional materials and process optimization using data-driven experimental design with well-calibrated uncertainty estimates. *Integrating Materials and Manufacturing Innovation* 6:207–217. <https://doi.org/10.1007/s40192-017-0098-z,05July>
- [57] Krige DG (1951) A statistical approach to some mine valuations and allied problems at the Witwatersrand, University of Witwatersrand: Master thesis.
- [58] Mathworks (2021) “Kernel (Covariance) Function Options,” [Online]. Available: <https://www.mathworks.com/help/stats/kernel-covariance-function-options.html>.
- [59] Firdous R and Stephan D (2019) “Influence of heat treatment and mechanical activation on reactivity of natural pozzolan for geopolymer synthesis,” 2 – 15 August, Eindhoven, The Netherlands
- [60] Kani EN and Allahverdi A (2009) “Effect of Chemical Composition on Basic Engineering Properties of Inorganic Polymeric Binder Based on Natural Pozzolan,” *Ceram. – Silik.* 53 (3), p. 195–204.
- [61] Tänzer R (2018) Alkalisch aktivierte Systeme auf Basis von Hüttensandmehlen (Dissertation). Technische Universität Berlin, Berlin
- [62] Firdous R and Stephan D (2019) “Factors effecting the properties of alkali activated natural pozzolan based geopolymers,” Aachen Germany, pp. 100–106.
- [63] CEN/TC 51, DIN EN 196–6:2010–05 Methods of testing cement - Part 6: Determination of fineness; German version EN 196–6:2010, 2010.
- [64] ASTM C311, Standard Test Methods for Sampling and Testing Fly Ash or Natural Pozzolans for Use in Portland-Cement Concrete, Vol 05.05 ed., Annual Book of ASTM Standards : ASTM International, 2001.
- [65] DIN 66126, Determination of specific surface area of disperse solids by the gas permeability technique - Blaine method, 2015.
- [66] EN 196–6, Methods of testing cement – Part 6: Determination of fineness, 2010.
- [67] DIN EN ISO 11885, Water quality - Determination of selected elements by inductively coupled plasma optical emission spectrometry (ICP-OES) (ISO 11885:2007); German version EN ISO 11885:2009, 2009.
- [68] EN 196–2, Methods of testing cement - Part 2: Chemical analysis of cement, 2005.
- [69] van der Maaten L, Hinton G (2008) Visualizing Data Using t-SNE. *J Mach Learn Res* 9:2579–2605
- [70] Mathworks (2021) “k-means clustering,” [Online]. Available: <https://www.mathworks.com/help/stats/kmeans.html>. [Accessed 4 April 2021].
- [71] Mathworks (2021) “t-Distributed Stochastic Neighbor Embedding,” [Online]. Available: <https://www.mathworks.com/help/stats/tsne.html>. [Accessed 4. April 2021].
- [72] Ortega PA, Wang JX, Rowland M, Genewein T, Kurth-Nelson Z, Pascanu R, Heess N, Veness J, Pritzel A, Sprechmann P, Jayakumar SM, McGrath T, Miller K, Azar M, Osband I and Rabinowitz (2019) “Meta-learning of Sequential Strategies,” CoRR.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.