Journal of Cleaner Production 418 (2023) 138221

Contents lists available at ScienceDirect

Journal of Cleaner Production

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

Data driven design of alkali-activated concrete using sequential learning

Christoph Völker [a,*], Benjami Moreno Torres [a], Tehseen Rug [b], Rafia Firdous [c], Ghezal Ahmad Jan Zia [a], Stefan Lüders [b], Horacio Lisdero Scaffino [b], Michael Höpler [b], Felix Böhmer [b], Matthias Pfaff [b], Dietmar Stephan [c], Sabine Kruschwitz [a,d]

[a] Bundesanstalt für Materialforschung und -prüfung, Unter den Eichen 87, 12205, Berlin, Germany
[b] iteratec GmbH, St.-Martin-Str. 114, 81669, Munich, Germany
[c] Technische Universität Berlin, Institute for Civil Engineering, Building Materials and Construction Chemistry, Gustav-Meyer-Allee 25, 13355, Berlin, Germany
[d] Technische Universität Berlin, Institute of Civil Engineering, Non-destructive Building Material Testing, Gustav-Meyer-Allee 25, 13355, Berlin, Germany

ARTICLE INFO

Handling editor: Zhen Leng

Original content: Green building materials: a new frontier in data-driven sustainable concrete design (Original data)

Keywords:
Sustainable building materials
Sequential learning
Alkali-activated building materials
Data-driven materials design

ABSTRACT

This paper presents a novel approach for developing sustainable building materials through Sequential Learning. Data sets with a total of 1367 formulations of different types of alkali-activated building materials, including fly ash and blast furnace slag-based concrete and their respective compressive strength and CO_2-footprint, were compiled from the literature to develop and evaluate this approach. Utilizing this data, a comprehensive computational study was undertaken to evaluate the efficacy of the proposed material design methodologies, simulating laboratory conditions reflective of real-world scenarios. The results indicate a significant reduction in development time and lower research costs enabled through predictions with machine learning. This work challenges common practices in data-driven materials development for building materials. Our results show, training data required for data-driven design may be much less than commonly suggested. Further, it is more important to establish a practical design framework than to choose more accurate models. This approach can be immediately implemented into practical applications and can be translated into significant advances in sustainable building materials development.

1. Introduction

The use of computer modeling has revolutionized the field of civil engineering, allowing engineers to design their most complex and ambitious designs from the comfort and affordability of their desks. This has brought a new era of design capabilities. At the same time, materials development has not kept pace with the digital revolution that has taken place in the construction industry since the 1960s. The inherent complexity and variability of building materials make them difficult to model and analyze, which has historically limited the development of new materials to labor-intensive laboratory testing. This is a major bottleneck in construction innovation, with far-reaching consequences due to the significant contribution to CO_2 emissions of materials in use today.

The highest emissions come from the production of Portland cement, which alone is responsible for around eight percent of anthropogenic emissions (Andrew, 2018). Unlike most materials, the emissions from cement production result from the burning process and the decarbonization of limestone. International environmental goals such as the European Green Deal (Comission, 2023a) aim to reduce these emissions to achieve carbon neutrality in Europe by 2050.

Newly developed cementitious binders low in calcium can reduce carbon dioxide emissions by 40–80 percent while maintaining comparable structural properties to traditional cement (Provis and van Deventer, 2014). However, developing these new binders, commonly known as alkali-activated materials (AAM) or geopolymers, is challenging. AAMs can be synthesized from a large variety of aluminosilicate feedstocks and a wide range of activator solution compositions to achieve the desired properties. This way, binders with varying properties suitable for various applications can be obtained. Unfortunately, this large variety makes it difficult to develop these new formulations in a feasible timeframe. Each raw material and binder could require batch-wise adjustment in the laboratory, which is very time-consuming given the complex nature of the underlying chemical reactions.

Another critical concern is that the future availability of well-established substitutes is uncertain. In particular, the quantities of fly

* Corresponding author.
E-mail address: christoph.voelker@bam.de (C. Völker).

https://doi.org/10.1016/j.jclepro.2023.138221
Received 28 February 2023; Received in revised form 19 June 2023; Accepted 21 July 2023
Available online 22 July 2023

ash will diminish in the foreseeable future because many countries are approaching the end of coal-fired power generation (Global Energy Monitor, 2013; Comission, 2023b). Additionally, due to the conversion of blast furnaces to hydrogen, the composition of granulated blast furnace slag substitutes will also differ significantly from the currently known (Mathieson et al., 2015). This could lead to production losses and setbacks and even undo the progress made so far in the ecological transformation of Portland cement.

In the future, alternatives including secondary raw materials, recycled materials and industrial by-products from various sources and processes could be used as binder components. The decision criteria for their application are not only given by material properties but also by material costs - monetary and environmental. These, in turn, depend on processing, source location, and material recovery. It has been shown that cost-critical components can be added and replaced with new ones in more complex mixed material systems, further increasing the overall economics of greener building materials (Gökçe et al., 2020). Nevertheless, using many new raw materials and their combinations, both with each other and with additives and possibly other primary resources, leads to an exponentially growing number of possible compositions. The synthesis and experimental characterization of these compounds would exceed the capacity of current lab-based experimental exploration.

Artificial Intelligence (AI)-based materials design methods solve the problem of optimally tuning formulations much more efficiently. Rather than relying solely on a time-consuming laboratory validation, optimization frameworks such as Sequential Learning (SL) and the closely related Bayesian Experimental Design (BED) (Lookman et al., 2019) create a data-driven feedback loop: From a variety of possible formulations, SL predicts the most promising candidate materials, which are then empirically validated in the laboratory. The predictive models are trained again with the empirical results from the laboratory to make better suggestions in the next round. Instead of calculating the parameters for an ideal formulation directly, Machine Learning (ML) models optimize formulations in an iterative cycle of predictions and validations.

Although the complexity of the building materials tuning does not yet permit a purely computer-aided design, it reduces the number of samples with undesirable properties. Formulations that do not provide valuable information or have little chance of success are discarded from the outset, resulting in more efficient use of limited lab resources.

Moreover, SL makes it possible to improve many material properties simultaneously and to include critical ecological and socioeconomic factors, e.g., carbon footprint, material costs, or resource availability as optimization criteria. In this way an ideal concrete precisely tailored to its purpose can be developed quickly while considering the given market conditions.

1.1. Novelty, scope, and research approach

This paper presents a large-scale investigation of the feasibility and practicality of developing alternative building materials with SL. The goal is to provide a guide on how SL optimization can be performed successfully against the constraint of a real-world laboratory.

Experiments for SL-driven concrete design were conducted to measure the performance in terms of the number of iterations (or development cycles) needed to find a material with desired properties. These experiments were repeated with randomized starting conditions and were assessed statistically.

Crucially, the materials selected by SL at each development cycle were not hypothetical. They were drawn directly from laboratory results previously reported in the literature (as summarized in Table 1), underscoring the real-world applicability of the findings. The data for these experiments were compiled from 49 sources, comprising 1367 formulations for alkali-activated concrete (compare Table 1). For each formulation, the resulting compressive strength from the laboratory is available - one of the critical properties of concrete. It's important to note that all data sets were carefully selected to have the same specimen shape, dimensions, and test age to ensure comparability. In addition, the data was enriched by including CO2 footprints based on literature values.

Two common predictive SL models were compared, Random Forest regression (RF) and Gaussian Process Regression (GPR) and the effectiveness of different pre-processing and model-tuning pipelines was investigated. Further, two search strategies for SL were compared, one relying solely on model predictions, to sample the next composition, and the other deliberately exploring uncertain compositions by considering both the predictions and their uncertainties.

To investigate the ideal size of initial training data sets in SL-based materials design all experiments were performed with 4 and 20 initial samples, respectively. The design target was, on the one hand, naively based on strength predictions only and, on the other hand, informed by the carbon footprint of the material.

The results provide hope for a much-needed opportunity to accelerate the "time to solution" for new climate-friendly building materials.

2. Points of departure

This chapter summarizes the literature and previous work and lays out the hypothesis, research questions and goals for exploration of this contribution.

2.1. Literature review

Previous studies have demonstrated that achieving high-quality building materials with acceptable ecological impact is possible. For instance, by varying base materials and optimizing synthesis pathways it is possible to improve both the performance and the ecological footprint of alternative materials such as alkali-activated concretes (AAC) (Gökçe et al., 2020; He et al., 2013; Provis et al., 2015). However, as Firdous et al. point out in (Firdous et al., 2022), the number of possible alkali-activated concrete formulations based on commonly used supplementary cementitious materials is beyond human capacity to test in the lab.

In addition, small changes in curing conditions, mixing methods, and material supply have a significant impact on the properties of the materials. Unfortunately, reliably predicting these properties is not feasible in general due to the underlying complexity of the composition and the production process. At the same time, recent environmental goals including the European Green Deal, encourage the use of alternative raw materials resources to achieve a green circular economy. Thus, alternative approaches to designing and manufacturing building materials are urgently sought after.

Because of the many different constituents contained in typical building materials and the subsequent variety in their properties, a computer-aided approach to material design was considered early (Addy and Koshla, 1986). Nonetheless, predictions based on empirical data were not possible at that time. The recent advent of ML methods in concrete science and increases in computing power have expanded the possibilities for material design. Naderpour et al. (2018) utilized Artificial Neural Networks (ANNs) to predict the compressive strength of environmentally friendly concrete mixtures. Chaabene et al. (2020) reviewed the effectiveness of different ML techniques, including ANNs, support vector machines, decision trees, and evolutionary algorithms, in predicting concrete properties. Li et al. (2022) provide a survey of applications and best practices in the field as of 2022, emphasizing the significance of data preparation, model validation, and their interpretation.

In the field of concrete based 3D printing, Sergis et al. (Sergis and Ouellet-Plamondon, 2022) employed a combination of ANNs, genetic algorithms, and Pareto optimization to find mixtures with optimal properties. Similarly, Golafshani et al. (2021) combined genetic

algorithms and decision tree ensembles to find concrete mixtures incorporating rubber that maintained compressive strength while reducing CO_2 emissions. Related techniques were used to predict novel concrete mixtures that aim to reduce costs (Zhang et al., 2020) and CO_2 footprint (Shobeiri et al., 2022). Further, a recent article uses generative design and genetic algorithms to optimize the topology and steel reinforcement of concrete beams leading to promising results with respect to saving costs and improving the environmental footprint (Alsakka et al., 2023). A detailed review on combining ML algorithms with evolutionary approaches is summarized by Song et al. (2022).

Although these frameworks propose new promising compositions, they suffer from a major drawback, as they rely on training ML algorithms on sufficiently large data sets (~1000 datapoints). Collecting this data is costly and time-consuming, especially in the field of building materials as elaborative experiments need to be performed to obtain the data. Furthermore, the varying composition of the precursor materials add additional optimization challenges. These challenges are also analysed in detail by Li et al. (2022).

SL has successfully adapted to complex new scenarios while getting by with relatively little training data (Lookman et al., 2019; Montoya et al., 2020). Ling et al. (2017) demonstrated the potential of SL in materials discovery across multiple domains, excluding concrete. Later, Rohr et al. (2020) benchmarked SL for applications in electrochemistry, comparing its performance to that of GPR, RF, and Linear Ensemble Regressors. They found that SL can significantly accelerate materials discovery when used properly, but caution against potentially negative effects if certain aspects of the model are not carefully chosen. In the context of building materials, Völker et al. (2021) investigated the use of SL for detecting alkali-activated binders and found it to be highly effective, requiring up to 60 times fewer data and processing more than three times as many features as conventional ML methods. This suggests superior performance in complex real-world scenarios. Von Rueden et al. (Von Rueden et al., 2021) pointed out that integrating prior knowledge in ML models can benefit various research areas. Völker et al. (2022) demonstrated that incorporating known material parameters as a knowledge-based loss term is valuable in guiding the search for environmentally friendly building materials.

Despite these promising results, there is no comprehensive understanding of the potential performance capabilities of data-driven building material design with sequential learning. The scarcity of available samples, the complex, diverse nature of starting materials, and the high cost of experimentation pose challenges for extensive data-driven work. The incomplete characterizability of these materials adds to the uncertainties, which may have hindered the adoption of data-driven approaches in the laboratory to date.

Despite these challenges, there is a significant opportunity to explore and demonstrate the potential of sequential learning for building materials in various scenarios. This knowledge gap presents an opportunity for further research to better comprehend the potential of using SL in construction.

2.2. Hypothesis, research question, goals for exploration

The central hypothesis of this study is that the configuration of the SL-driven materials design (e.g., in terms of model selection, hyperparameter settings, and selected strategy) has a significant impact on the expected performance. To probe this hypothesis, we aim to answer the following research questions that will help us to better understand the potential and limitations of SL in this context:

1. What is the impact of:
   a. Limited training data points commonly available for concrete testing on the effectiveness of SL?
   b. Incorporating ecological information, in particular the CO2 footprint, on the performance of SL in finding optimal AACs?
   c. Using classical GPR models compared to RF models for a given scenario, and can automated pipelines for model optimization and feature selection further enhance SL performance?
   d. Exploratory, uncertainty-based techniques compared to point estimator predictions in finding optimal AACs, and under what circumstances is one approach more effective than the others?
2. How effectively can SL accelerate the search for optimal AAC?
3. Considering that only a small portion of the training data is labeled at the end of a material design run, how well do the different models generalize to the remaining data points?

Recognizing that the performance of SL algorithms in real-world applications can be significantly influenced by the specific characteristics of the underlying data set and the complexity and unpredictability of the optimization target (as discussed in (Kim et al., 2020)), this study aims to provide a comprehensive guide for the practical application of SL in the field of materials design.

The primary objectives of this study are:

- To demonstrate the potential of SL in identifying optimal AACs, thereby providing a concrete example of how this approach can accelerate the development of sustainable building materials.
- To offer clear and actionable strategies that can enhance the performance of SL algorithms, even under challenging conditions such as small training samples, high-dimensional input data, and discontinuous batches of material.
- To provide insights into the effectiveness of SL in the context of materials design, thereby improving understanding of the role of configuration and other factors in the performance of these algorithms.

By achieving these objectives, this study aims not only to advance the field of concrete design but also to provide a roadmap for other researchers and practitioners. The ultimate goal is to facilitate the adoption of more effective data-driven materials design, thereby accelerating the development of sustainable building materials and contributing to a more sustainable future.

3. Research method

Traditional material development workflows typically start by developing the formulation (i.e., prescriptive-based material compositions) to achieve the desired properties – for instance a required compressive strength - that can be validated in the laboratory. However, predicting the material composition can be challenging for various reasons (Gallet et al., 2022). First, formulations often need to be better-conditioned, i.e., small changes in composition can significantly affect the resulting material properties. Second, numerous non-unique solutions exist where completely different compositions with different amounts of constituents result in the same target properties. Finally, the number of constituents in the solution is not fixed, making a closed-form solution unattainable.

The Inverse Design (ID) approach is a viable method for addressing these challenges. It involves screening a set of given formulations and iteratively identifying optimal compositions based on how well their predicted properties align with the target requirements. This approach turns the material development process around by optimizing the predicted target property rather than the composition. As a result, ID has the potential to uncover novel materials that meet the desired specifications but would not be apparent through traditional design workflows.

This chapter describes the implementation of the ID approach using SL and highlights the differences in benchmarking SL and the underlying models. The data on which this work is based is also presented.

3.1. Description of the SL-guided materials design tasks

SL-based ID uses a prediction model to guide the selection and evaluation of material formulations to identify optimal material properties. This iteratively improves the properties of the material by following the steps in Fig. 1.

1. Creating a materials search space or design space (DS) that contains the formulations of viable materials that can be considered for the desired outcome.
2. Training an ML model to predict the properties of materials based on their composition and structure using the initial training data. Viable models include GPR for BED, and RF or any other suitable regression algorithm for SL.
3. Selecting additional samples for characterization using ML prediction weights by acquisition function, also known as a sample utility function.
4. Experimentally validating the selected samples in the laboratory.
5. If the criteria are not met, updating the ML model with the new data, and repeating the process allowing for continuous improvement in the known material's properties.
6. If the criteria are met, the design task is finished.

The DS should be extensive and inclusive yet limited to material formulations that are physically feasible to produce. The initial training data is obtained by selecting and characterizing a representative subset of the entire material space, serving as the starting point for the ML model.

The design process is guided towards target criteria at each iteration using a utility function. This function assigns weights to candidate materials based on their predicted properties and any available prior information. The closer the predicted result aligns with the desired value, the higher the weight assigned to the material. Additionally, the utility function may consider prediction uncertainty, which is crucial in discovering new relationships and supporting experimentation decisions. As Reyes et al. (Reyes and Powell, 2020) aptly put it, if the outcome of an experiment is already certain, there is no reason to conduct it, and an experiment is more useful if the uncertainty of the outcome is significant. All three factors (prediction, uncertainty, and a-priori information) are included in the utility function in equation (1).

$$u_i = \mu_i + w * \sigma_i - A_i \tag{1}$$

where μ_i corresponds to the normalized predicted material property (e.g., compressive strength), $w * \sigma_i$ is the weighted uncertainties of the i-th prediction and A_i are the normalized a-priori information (e.g., carbon emissions). This formula applies to Multi-Objective Optimization (MOO) scenarios. It calculates the joint performance based on the normalized sum of targets, which includes predicted material properties and a priori information. Normalizing the targets to a common scale enables fair comparison of otherwise disparate quantities during the optimization process. Objectives that are to be minimized, such as carbon emissions, are given a negative sign to reflect the preference for materials with lower CO2 to have higher utility. In Single Objective Optimization (SOO), only the predicted material property is sought, resulting in A_i being omitted.

The prioritization of the next candidate x_{n+1} is done by choosing the maximum value according to equation (2).

$$x_{n+1} = argmax(u) \tag{2}$$

This systematic exploration of the space of possible material compositions allows directly identifying materials with the desired properties. There are two predominant configuration strategies: exploratory and exploitative. The exploratory strategy is characterized by a high weight of uncertainty, which is intentionally leveraged to acquire new information rapidly. In contrast, the exploitative strategy is focused on utilizing the predictions of the SL model to attain the desired outcome with a higher probability of success, which necessitates robust predictive performance. However, since in typical real-world scenarios only a few samples are available when starting a SL-guided feedback loop, predictive performance is expected to be limited. Thus, the choice between the two strategies often involves balancing the trade-off between acquiring new information and exploiting the model's predictions. Further information on this approach can be found in (Lookman et al., 2019) and (Reyes and Powell, 2020).

In the context of this study, two distinct settings were employed to navigate the trade-off between exploration and exploitation. A weight of zero puts the model in an 'exploitative' mode, relying solely on point estimates of predicted material properties, without considering prediction uncertainty. This strategy aimed to utilize the SL model's predictions to achieve the desired outcome.

On the other hand, a weight of two put the model in an 'exploratory' mode. In this mode, the model leveraged prediction uncertainty to rapidly acquire new information. The next candidate was selected according to the upper bound of its utility probability estimate. This means that the model prioritized candidates with the highest possible utility value, considering both their predicted properties and the associated uncertainties. In essence, this strategy favored candidates that not only showed high potential for desired properties but also carried significant uncertainty, thereby offering the greatest potential for learning and discovery. The choice of '2' as the weight, while arbitrary, underscores the concept of incorporating predictive uncertainty into decision-making, representing a deliberate effort to explore the material composition space more broadly and identify materials with desired properties more rapidly.

3.2. Benchmarking SL

Sequential Learning's (SL) effectiveness is best demonstrated through simulated experiments where all data point outcomes are known. This approach enables multiple explorations of a given design space under randomized initial conditions, thereby providing a statistical measure of the exploration performance in terms of the required duration. As depicted in Fig. 2, the benchmarking workflow operates as follows.

Initially, a subset of the available data is provided to the SL algorithm (marked in green in Fig. 2). With each development cycle, the algorithm selects and incorporates the data point predicted to be the most promising (marked with red arrows in Fig. 2) according to the utility estimate from equation (1). The optimization's success is gauged by the number of development cycles required to identify materials with the desired target properties (marked in blue in Fig. 2).

The benchmarking outcome is the distribution of the number of SL cycles required (compare Fig. 2, bottom), which can vary based on the experiment's initial conditions. This distribution is evaluated statistically, for instance with the 90th percentile serving as an indicator. This percentile corresponds to the point where 90% of the SL runs achieved the target, indicating a high probability that the target can be reached within the specified number of SL cycles.

SL is typically compared against a random draw (RD) as a baseline benchmark. In RD, experiments are carried out without a strategy or model, and the probability of success can be estimated using the hypergeometric distribution. More details on benchmarking SL can be found in (Völker et al., 2021).

3.3. Benchmarking the prediction model

In addition, the predictive performance for the different model and pipeline combinations are evaluated using the coefficient of determination (R-squared) (Steel and Torrie, 1960). This metric measures the performance of the model by comparing the predicted and actual values of a dataset. The dataset is divided into two parts: a training set and a test

set. The training set is used to build the model, while the test set is used to evaluate its performance on unseen data.

Two scenarios are compared: (1) a baseline performance obtained using a random 70/30 training/test-split of the data, and (2) the model's prediction on the remaining data points at the end of each SL run. In scenario (1), the training set is split into 25 randomly selected folds, a common practice that assumes Independence and Identically Distributed (IID) data. This shows the general predictive limits imposed by the underlying data. In scenario (2), the samples are drawn during the optimization process, and the performance of the model is measured by calculating R-squared for its predictions on the remaining data. To ensure a reliable performance assessment the process is repeated 25 times. This accounts for the variability of the SL sampling and the potential for overfitting, which can result from different training sample selections. The final performance of the model on each dataset is calculated as the average of the obtained R-squared values.

3.4. Description of the data

The data sets used in this work are taken from the compilation by Xie et al. (2020). They contain formulations for alkali-activated concretes and the corresponding compressive strengths. Nine data sets of the same specimen shape, dimensions and test age were extracted. In addition, these data sets were enriched with CO_2 footprints calculated from CO_2 values of constituents obtained from the literature. Both compressive strength and CO_2 emissions are critical to selecting concrete for various applications.

Table 1 gives an overview of the data sets showing the original references, the different compositions, the binder type, the number of mixes per study, the shape and dimensions of the specimens and their testing age. Each data set contains data from an average of 10 papers and 152 formulations. This results in over 1300 materials, each described by 28–43 descriptive features, the corresponding compressive strength values from the laboratory, and the calculated carbon footprint. The

Table 1
Description of data sets and 95% quantile target thresholds.

Input data description							95% Target SOO		95% Target MOO	
Subset	Nr of distinct REFs	[Reference] -Materials-(Nr. Of mixes)	Shape, dimensions (mm)	Testing age (days)	Nr. Of formulations	Nr. Of formulation parameters	Strength (MPa)	CO_2 (kg/m^3)	Strength (MPa)	CO_2 (kg/m^3)
1	7	(Memon et al., 2011)-FA-(8); (Memon et al., 2013)-FA-(4); (Vora and Dave, 2013)-FA-(17); (Omar et al., 2015)-FA-(6); (Vinai et al., 2016)-FA/GGBFS-(18); (Rafeet et al., 2017)-FA/GGBFS-(39); (Fang et al., 2018)-FA/GGBFS-(13)	Cube, 100^3	1	105	39	50.3	282.3	46.0	224.5
2	7	(Aliabdo et al., 2016a)-FA/OPC-(17); (Aliabdo et al., 2016b)-FA-(14); (Haddad and Alshbuol, 2016)-NP-(8); (Shehab et al., 2016)-FA/OPC-(16); (Chithambaram et al., 2018)-FA-(40); (Sun et al., 2018)-FA/GGBFS/OPC-(4); (Aliabdo et al., 2019)-GGBFS-(27)	Cube, 100^3	7	126	36	63.0	211.5	61.1	187.0
3	17	(Memon et al., 2011)-FA-(8); (Jeyasehar et al., 2013)-FA-(12); (Kumar et al., 2013)-FA/GGBFS-(4); (Barnard, 2014)-FA/GGBFS-(9); (Parthiban and Mohan, 2014)-GGBFS-(12); (Vinai et al., 2016)-FA-(24); (Parthiban and Vaithianathan, 2015)-GGBFS/MK-(9); (Abhilash et al., 2016)-FA/GGBFS-(3); (Haddad and Alshbuol, 2016)-NP-(17); (Muthadhi et al., 2016)-FA-(5); (Vinai et al., 2016)-PFA/GGBFS-(18); (Rafeet et al., 2017)-FA/GGBFS-(39); (Honggen et al., 2017) -FA-(6); (Fang et al., 2018)-FA/GGBFS-(13); (Sun et al., 2018)-FA/GGBFS/OPC-(8)	Cube, 100^3	7	188	43	56.5	137.1	54.3	137.1
4	9	(Aliabdo et al., 2016a)-FA/OPC-(17); (Aliabdo et al., 2016b)-FA-(14); (Haddad and Alshbuol, 2016)-NP-(8); (Shehab et al., 2016)-FA/OPC-(16); (Shinde and Kadam, 2016)-FA-(9); (Manickavasagam and Mohankumar, 2017)-FA-(4); (Chithambaram et al., 2018)-FA-(40); (Sun et al., 2018)-FA/GGBFS-(4); (Aliabdo et al., 2019)-GGBFS-(27)	Cube, 100^3	28	139	37	68.8	198.1	64.2	187.0
5	20	(Memon et al., 2011)-FA-(8); (Kumar et al., 2013)-FA/GGBFS-(7); (Memon et al., 2013)-FA-(4); (Barnard, 2014)-FA/GGBFS-(41); (Parthiban and Mohan, 2014)-GGBFS-(12); (Aravindan et al., 2015)-FA/GGBFS-(4); (Omar et al., 2015)-FA-(24); (Parthiban and Vaithianathan, 2015)-GGBFS/MK-(12); (Abhilash et al., 2016)-FA/GGBFS-(3); (Haddad and Alshbuol, 2016)-NP-(17); (Muthadhi et al., 2016)-FA-(5); (Pavithra	Cube, 100^3	28	250	43	68.0	129.4	66.9	129.4

(continued on next page)

Table 1 (*continued*)

Input data description							95% Target SOO		95% Target MOO	
Subset	Nr of distinct REFs	[Reference] -Materials-(Nr. Of mixes)	Shape, dimensions (mm)	Testing age (days)	Nr. Of formulations	Nr. Of formulation parameters	Strength (MPa)	CO_2 (kg/m^3)	Strength (MPa)	CO_2 (kg/m^3)
		et al., 2016)-FA-(5); (Vinai et al., 2016)-PFA/GGBFS-(17); (Manickavasagam and Mohankumar, 2017)-FA-(4); (Rafeet et al., 2017)-FA/GGBFS-(39); (Honggen et al., 2017)-FA-(6); (Al-Tais and Annapurna, 2018)-VA/FA/RHA/GGBFS-(16); (Fang et al., 2018)-FA/GGBFS-(13); (Reddy et al., 2018)-FA/GGBFS-(5); (Sun et al., 2018)-FA/GGBFS/OPC-(8)								
6	4	(Bidwe and Hamane, 2015)-FA-(9); (Rao and Rao, 2018)-FA/GGBFS-(198); (K and Babu, 2018)-FA/GGBFS-(36); (Rai et al., 2018)-FA-(31)	Cube, 150^3	28	274	32	59.9	174.5	33.9	133.3
7	1	(Hardjito and Rangan, 2005)-FA-(105)	Cylinder, 100 × 200	7	105	28	76.0	173.2	55.0	106.6
8	9	(Olivia et al., 2011)-FA-(8); (Adam, 2009)-FA/GGBFS-(6); (Jeyasehar et al., 2013)-FA-(12); (Nath and Sarker, 2014)-FA/GGBFS-(8); (Albitar et al., 2015)-FA-(3); (Nath and Sarker, 2015)-FA/OPC-(7); (Topark-Ngarm et al., 2014)-FA-(12); (Hadi et al., 2017)-GGBFS/FA/MK/SF-(28); (Hongen et al., 2017) -FA/OPC-(6)	Cylinder, 100 × 200	7	90	39	56.1	179.0	37.0	125.5
9	14	(Olivia et al., 2011)-FA-(8); (Adam, 2009)-FA/GGBFS-(6); (Pan et al., 2011)-FA-(3); (Sarker et al., 2013)-FA-(3); (Nath and Sarker, 2014)-FA/GGBFS-(8); (Albitar et al., 2015)-FA-(3); (Nath and Sarker, 2015)-FA/OPC-(7); (Topark-Ngarm et al., 2014)-FA-(12); (Wardhono, 2015)-FA/GGBFS-(2); (Noushini and Castel, 2016)-FA/MK/GGBFS-(13); (Hadi et al., 2017)-GGBFS/FA/MK/SF-(9); (Nath and Sarker, 2017a)-FA-(5); (Nath and Sarker, 2017b)-FA/GGBFS-(5); (Hongen et al., 2017)-FA/OPC-(6)	Cylinder, 100 × 200	28	90	38	62.3	274.7	55.8	158.2

features can be grouped into the following six categories:

1) The oxide composition of the binder powder is given in weight percent, which provides information on the chemical reactivity of the constituents. If only one binder is used in the concrete mix, the values are stated directly in the work. If more than one binder is used, the weighted ratio is calculated for each molecule.
2) The weights are given for each type of precursor, namely fly ash (FA), ground granulated blast furnace slag (GGBFS), metakaolin (MK), natural pozzolan (NP), and OPC.
3) The weight of coarse and fine aggregates per cubic meter of concrete is given.
4) The alkali solution is described by its chemical composition, which is characterized by a minimum of 8 and a maximum of 14 features. Specifically, two types of alkalis are described: sodium hydroxide (NaOH) and sodium silicate (Na_2SiO_3). The former is based on the dissolution of NaOH pellets in water, while the latter is commercially available and is characterized by its content of Na_2O, SiO_2 and H_2O.
5) The amount of additional water and/or superplasticizer is indicated. This includes the total amount of water in the concrete mix, i.e., the sum of the additional water and the water contained in the alkalis.
6) The curing process is described. The quantitative characteristics of the source are retained, and the non-quantitative characteristics are discarded, except for "curing at ambient temperature". In this case, 25 °C is considered as the ambient temperature for curing.

The CO_2 footprint has been calculated for each mix using the approach of (Alsalman et al., 2021) and has been extended with a linear regression of the CO_2 footprint due to hot curation from (Yang et al., 2017). Both factors are summarized according to the following heuristic equation:

$$CO_2\ emissions\ \left(\frac{t}{m^3}\right) = \sum_i w_i \cdot m_i + (0.6417 \cdot T - 16.0417) \cdot t \quad (3)$$

where:

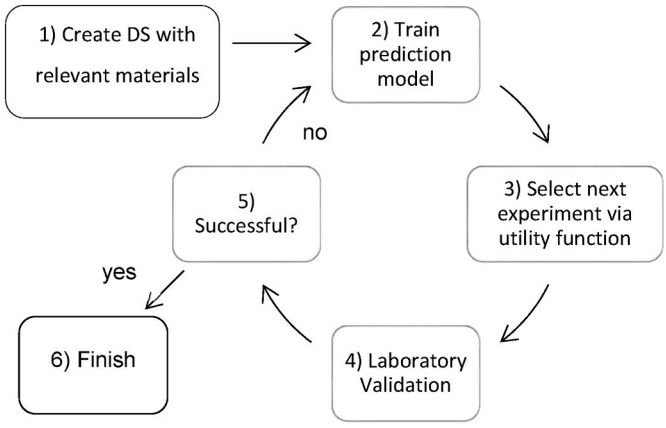

Fig. 1. Steps of the sequential learning inverse design loop.

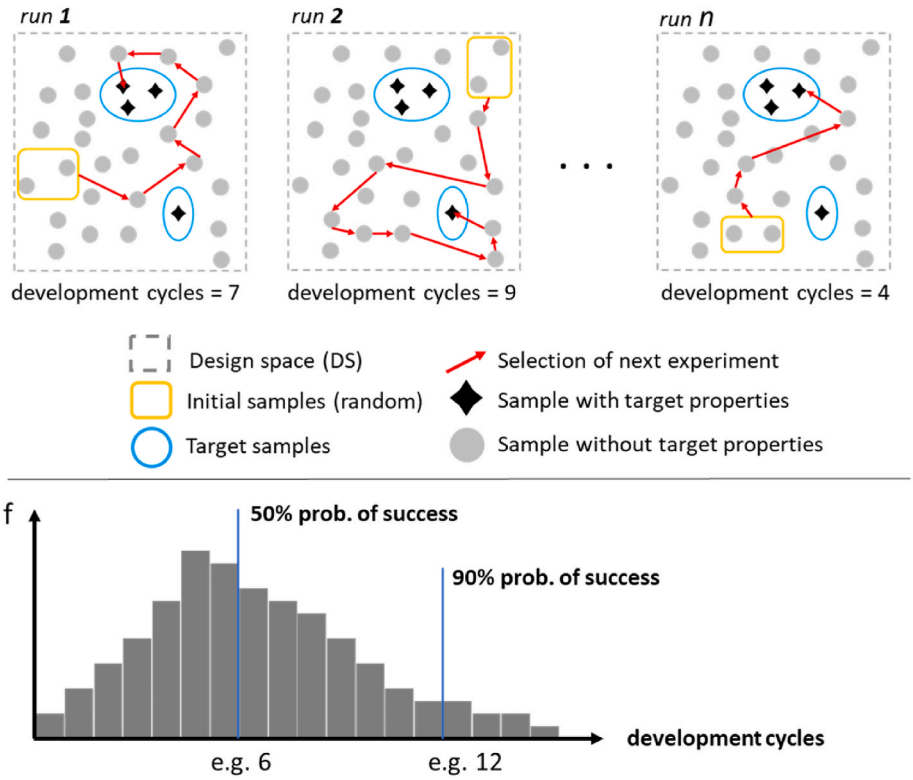

Fig. 2. Workflow of SL-benchmarking experiments; Top: SL iteratively samples a given DS starting with n randomized initial samples (marked in yellow) to find target materials (marked in green); Bottom: Histogram of required samples with 50% and 90% success probability. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2
CO_2 footprint for the precursor materials in (t CO_2)/(t precursor material).

	OPC	Fly Ash	GGBFS	Silica Fume	Metakaolin	Aggregates	Superplasticizer	NaOH Dry	Sodium silicate solution	Sodium silicate dry
w (t/t)	0.84	0.004	0.052	0.014	0.33	0.0048	1.88	1.915	0.36	1.222

w_i CO_2 emissions (t) to produce 1 t of the mix component i
m_i Mass of a mix component i in t/m^3 of fresh concrete
T Heat curing temperature if higher than 25 °C; otherwise, t is set to zero
t Curing time in days

The first part of the equation sums the carbon footprint w_i weighted with the mass of the individual precursor materials m_i (Alsalman et al., 2021). provide the values shown in Table 2.

The second part of equation (3) represents the cumulative emissions from heat treatment. It is modeled as a linear function that contains the curing temperature **T** in °C and duration **t** in days and is applied only to materials with the curing temperature higher than 25 °C, otherwise it is set to zero (assuming that ambient curing does not contribute to carbon emissions).

Table 3
Parameters and values used in the SL material discovery approach.

	Parameter	Value
1)	Optimization target	Single objective: compressive strength f_c (max) Multi objective: compressive strength f_c (max) & CO_2 footprint (min)
2)	Target quantile	95%
3)	Model	Gaussian Process Regression Random Forest Regression
4)	Pipeline	Vanilla Vanilla + dimensional reduction with PCA Simple grid search
5)	Strategy	Exploitation ($w_\sigma = 0$) Exploration ($w_\sigma = 2$)
6)	Initial sample size	4, 20 (batch size = 1)

4. Experimental program

The scenarios examined are described in detail below and are summarized in Table 3.

First, two optimization scenarios for material design are considered. SOO is performed first, in which materials with high compressive strength are sought. This optimization represents a laboratory-driven approach, where only predicted material strength is considered for decision-making. It is the most common case in the material optimization literature.

For the second optimization target, the best compromise between CO_2 footprint and strength is sought. This optimization scenario represents a more practical-oriented approach, in which the suitability of a material for a particular application may depend on a whole range of properties that do not have to come exclusively from the laboratory.

Second, the ID method used in this study establishes a target performance threshold using a quantile measure. Specifically, a 95% target quantile is selected to represent the top 5% of material properties with the highest performance level. In other words, the optimization process is completed once a material is found that exceeds the performance level of 95% of the data set.

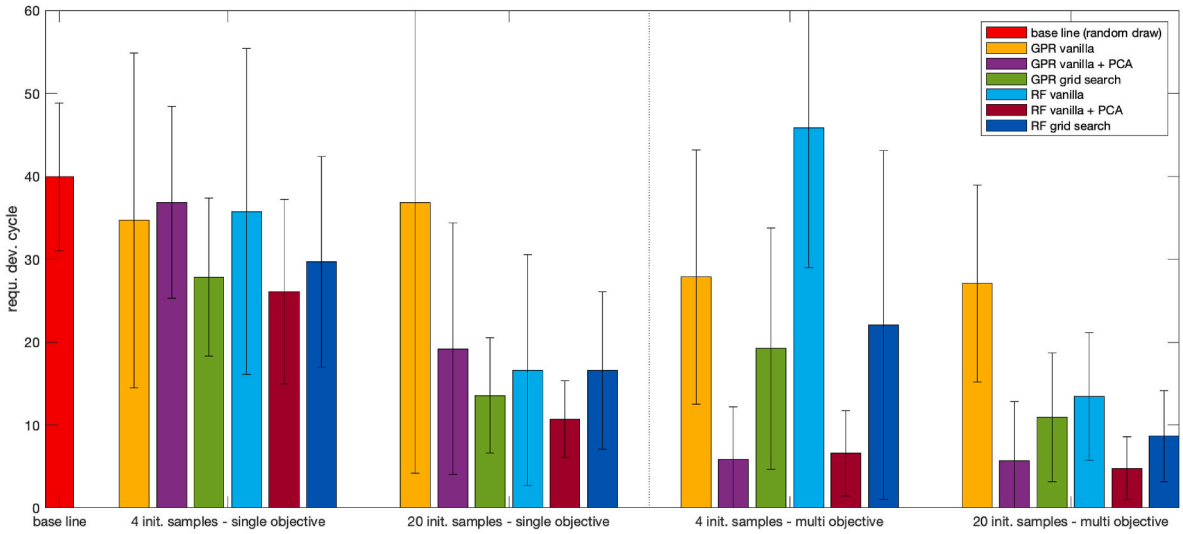

Fig. 3. Performance benchmark of SL algorithms in the exploitative scenario, measured by the number of development cycles required to achieve a 90% success probability. The results represent the mean performance across nine different data sets. The left side illustrates the results for single objective optimization, while the right side presents the outcomes for multi-objective optimization.

The material properties corresponding to this 95% performance threshold are summarized in Table 1 for each data set. To show the CO_2 savings potential, the corresponding CO_2 values have also been listed for the SOO, although they are not considered in the actual optimization. The use of the upper 95% quantile as the target criterion is motivated by the fact that the data comes from various references and may contain outliers, which could bias the estimate of the maximum strength. By using the upper 95% quantile instead of the maximum, the target criterion is based on a more robust and conservative estimate of the strength distribution, which is less susceptible to the effects of outliers.

Third, the predictive Models GPR and RF were used. Both techniques are considered non-parametric, meaning that they do not make assumptions about the functional form of the data and can flexibly adapt to complex patterns in the data (Kotlar et al., 2019). One key difference between GPR and RF is the way they model uncertainty. GPR models uncertainty using a Bayesian framework, in which a probability distribution is assigned to each predicted value. This allows the model to not only make predictions, but also to provide a measure of uncertainty for each prediction. In contrast, RF does not explicitly model uncertainty and does not provide a measure of uncertainty for their predictions.

Table 4
Best design decisions for SL configurations and potential improvement as an average of the required development cycles over nine collected data sets.

Decision	Best decision (requ. Dev. cycle)	Alternative (requ. Dev. cycle)	Improvement (requ. Dev. cycle)
Target	MOO (13.8)	SOO (22.2)	8.4
Model	GPR (17.7)	RF (18.3)	0.6
Pipeline	Vanilla + PCA (13.7)	Vanilla (23.5)	9.8
		Grid Search (16.7)	3
Strategy	Exploration (15)	Exploitation (20.9)	5.9
Initial sample size	20 (13.9)	4 (22)	8.1

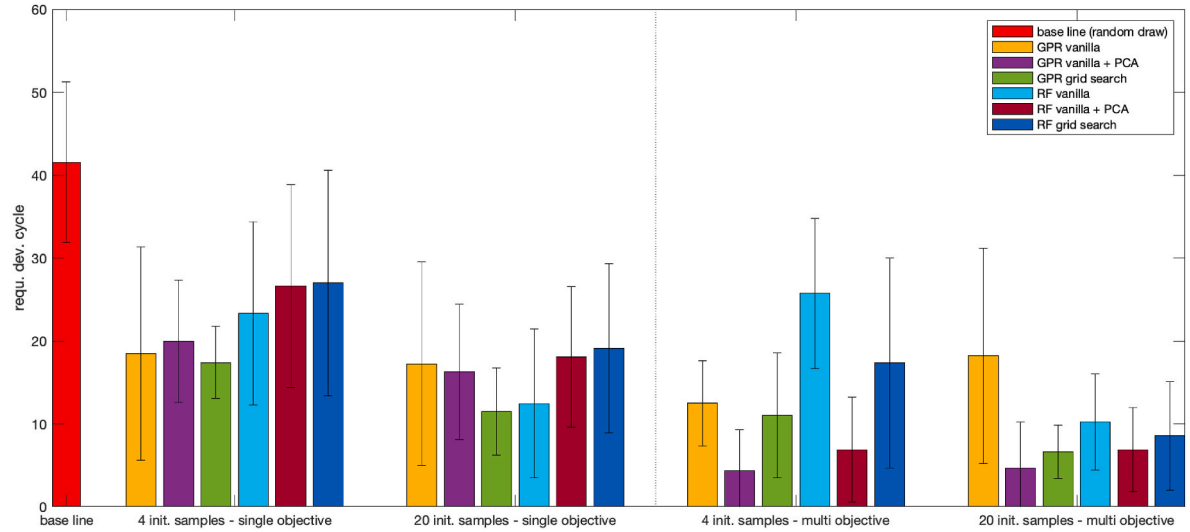

Fig. 4. Performance benchmark of SL algorithms in the explorative scenario, measured by the number of development cycles required to achieve a 90% success probability. The results represent the mean performance across nine different data sets. The left side illustrates the results for single objective optimization, while the right side presents the outcomes for multi-objective optimization.

Fortunately, subsampling methods such as jackknife bootstrapping enable estimating the uncertainty of a ML model's predictions (Wager et al., 2014).

In terms of performance, GPR tends to be more accurate than RF when the data is smooth and well-behaved. In contrast, RF can be more efficient when the data is large or has complex patterns.

Overall, the choice between GPR and RF depends on the specific characteristics of the data and the goals of the prediction task. In some cases, one method may be superior to the other. In other cases, both methods may be suitable, and the decision will depend on the specific requirements of the application.

GPR is described in greater detail in (Lookman et al., 2019) and the RF algorithm with uncertainty estimates has been introduced in (Ling et al., 2017). For this work the scikit-learn implementation of GPR (Scikit Learn, 2022) and the Lolo RF (Bousquet, 2017) have been implemented.

Fourth, three different model pipelines are compared: vanilla versions of the algorithms, models evaluated after dimensional reduction using Principal Component Analysis (PCA), and models tuned using a simple parameter grid and automated feature selection.

The vanilla versions of the algorithms are the simplest and most straightforward implementation of the regression models. These versions do not incorporate any additional techniques or modifications to improve the model's performance. As a result, they may not perform as well as more sophisticated approaches.

The second approach is dimensionality reduction using PCA. This technique aims to reduce the dimensionality of the data and remove redundant or irrelevant information. It is a common pre-processing technique mentioned in the literature on the implementation of ML (e.g., in (Ye and Wang, 2023)). In this work the dimensions were reduced such that 99% of variance of the data is kept.

The third approach involves tuning the model using a simple hyperparameters grid and automated feature selection. The hyperparameter grid contained two different kernels (isotropic and anisotropic) for GPR and two different tree depths in the case of RF. This was combined with a non-floating forward feature selector (Raschka, 2018) that kept the eight most promising features. This selector adds features one-by-one and uses an ML model to make predictions using the given feature set at each step. The optimal configuration was chosen based on four-fold cross-validation performance. In each iteration of sequential learning, the described parameter scan was performed and the best parameters for that iteration were selected in terms of best model fit based on the R2 score metric.

Fifth, an exploiting and an exploratory strategy were investigated by setting the weight of uncertainty in the utility function (compare equation (1)) to zero or two, respectively. Note that the equation reduces to a simple constant displacement of the predicted compressive strength minus the CO2 emissions when $w = 0$. In this 'exploitative' mode, the model relies solely on the point estimates of the predicted properties of the materials, without considering the uncertainty of these predictions. This approach is focused on utilizing the predictions of the SL model to attain the desired outcome with a higher probability of success.

Analogously, the case $w = 2$ adds approximately two standard deviations of variable displacement to the predicted compressive strength, which highlights how much the u_i could differ from the predicted μ_i. In this mode, the model intentionally leverages the uncertainty of the predictions to acquire new information rapidly. This strategy thus explores the space more eagerly, earning the name 'exploratory'.

Six, the initial training set size has been varied between 4 and 20 samples. While these numbers are arbitrary, they were chosen to represent two distinct scenarios that might occur in practical applications. A smaller set of 4 samples simulates a situation where laboratory capacity or resources are limited, often the case when developing entirely new materials. Conversely, a larger set of 20 samples reflects scenarios where more data is readily available, such as when fine-tuning well-known materials for novel applications like high-performance or ecologically engineered materials. After the initial training set size has been established, in each subsequent round, one sample is drawn (batch size = 1) based on the utility described above..

All calculations were conducted using the Benchmarking module of the Sequential Learning Benchmarking App (Völker and Völker, 2022). The purpose of this module is to test Sequential Learning in various configurations on known datasets and compare their results to those obtained using random draws as a baseline.

Finally, the predictive power of each algorithm-pipeline combination has been assessed using the R-squared benchmark. A baseline benchmark was created using 70% randomly sampled data points as a training set and the remaining 30% as a test set. This represents a scenario where the model is trained on a large set and provides a baseline benchmarking value to compare performance to what would ideally be possible.

The benchmarks for the SL with exploit and SL with explore strategy were created using only the data sampled at the end of the material optimization to assess the prediction performance of models on materials not considered in the design process.

To reduce the effect of outliers in the sampling - both during the optimization and the random sampling - the average performance is estimated over 25 repetitions for each data set. The final benchmark is the median performance over the 9 data sets.

Table 5
Ideal SL configuration results for different scenarios in terms of algorithm, pipeline and strategy.

Scenario	4 init. Sample (requ. Dev. cycle.)	20 init. Sample (requ. Dev. cycle.)
SOO	Explore GPR grid search (18.5)	Explore GPR grid search (10.7)
MOO	Explore GPR + PCA (4.4)	Explore GPR + PCA or Exploit RF + PCA (4.7)

Table 6
Benchmarking model performance for different algorithm pipeline combinations as the median R-squared over the mean performance from 25 randomized runs on 9 datasets.

Algorithm	Baseline performance	SL with exploit strategy	SL with explore strategy
GPR vanilla	0.05	0.12	0.27
GPR vanilla + PCA	0.35	−0.02	0.05
GPR grid search	0.08	−0.17	0.08
RF vanilla	0.79	0.36	0.42
RF vanilla + PCA	0.70	0.01	0.06
RF grid search	0.69	0.15	0.22

5. Results

A total of 10,800 experiments were conducted to evaluate the performance of two prediction algorithms (RF and GPR) and three pipelines (vanilla, PCA, grid) for two different targets (SOO and MOO) using two different initial training dataset sizes (4 and 20) and an exploitative or exploratory optimization strategy. These experiments were repeated 25 times under randomized conditions on nine different datasets, resulting in 225 experiments per bar shown in the figures below for each strategy. The results are shown as the number of development cycles required to reach the target within 90% of the runs (90% quantile of results), as an average across all data sets. A lower value indicates better performance. The results are compared against a RD as the baseline benchmark.

Figs. 3 and 4 present the results, categorized by the applied strategy: Fig. 3 for the exploitative strategy (weight of uncertainty = 0) and Fig. 4 for the explorative strategy (weight of uncertainty = 2). Each figure

represents 18 distinct experiment configurations per scenario, derived from two possible initial sample sizes (4 or 20), two ML models (GPR or RF), three pipeline implementations (Vanilla, Vanilla + PCA, or a simple grid search), and two optimization targets (SSO or MMO).

Table 4 shows the results for all possible scenario combinations and focuses on the impact of different SL configurations on the required number of development cycles based on a single factor considered. The table shows the optimal setting in the first column and the non-optimal alternatives in the following columns. The value refers to the changes in a particular factor while keeping all other factors constant. In other words, the table shows how much the required number of development cycles changes when only this factor is varied and all other factors remain the same.

With an average of 18 development cycles, the use of SL methods was generally more effective than the RD, which required 44 cycles to identify the target materials in the DS, which contained an average of 152 formulations.

The size of the initial data set was found to affect the efficiency of the optimization process. Larger data sets result in faster optimization, requiring an average of 13.9 development cycles, while smaller data sets require 22 cycles. However, if only four data points are initially available, the additional acquisition of 16 data points to improve performance may not justify the savings of a mere eight experiments. Overall, the number of initial samples does not seem to affect which model-pipeline combination produces the best results, i.e., there is no approach that is particularly well suited for small initial data sets but not for larger initial data sets. Performance appears to be invariant in this regard.

Using the carbon footprint information in the MOO significantly enhanced performance by reducing the average number of development cycles to 13.8, compared to 22.2 cycles for the SOO, where only laboratory variables were used. This reduced the CO_2 footprint by more than 21% (from an average of 195 kg/m^3 to 155 kg/m^3), while the compressive strength decreased by only 12% from 62.3 MPa to 54.8 MPa (compare Table 1).

The optimization strategy used also impacted performance, with an exploratory approach resulting in an average of 15 development cycles compared to 20.9 cycles for the exploit strategy. This difference was more pronounced for the GPR method (8.9 cycles improvement) than for the RF method (2.9 cycles improvement).

Regarding algorithmic choice, both GPR and RF performed similarly, with a slight advantage for GPR (17.7 development cycles compared to 18.3 for RF). The choice of the pipeline configuration seems to have the strongest average impact on the SL performance.

It was found that dimensional reduction using PCA was the most effective, with an average of 13.7 development cycles compared to 23.5 cycles for the vanilla implementation and 16.7 cycles for the grid search. However, the advantage of PCA seems much more pronounced in the MOO scenario - especially for exploitative strategy. In the SOO the grid-search brings the greatest improvement for GPR. This is in line with what previous studies have reported (Liang et al., 2021). RF benefits from the PCA pipeline in the exploit scenario and the vanilla implementation in the explore scenario. The latter is likely caused by unrealistically small uncertainties of the predictions in the dimensionally reduced space which is a disadvantage for the explorative approach.

The most challenging scenario was SOO with four initial training samples, which required an average of 18.5 development cycles even with the best available algorithms. In contrast, the ecologically informed MOO scenario with 4 initial samples was much more feasible, requiring an average of only four development cycles. The best configurations for the SOO and MOO scenarios with either four or twenty initial training samples are summarized in Table 5.

Finally, the model performance has been assessed (see Table 6). The best baseline result in terms of sheer predictive performance was achieved with the vanilla RF base model, closely followed by the RF + grid search model. While an R-squared of less than 0.8 is not an outstanding performance, it can be explained by the inconsistent data from globally distributed laboratories and by the relatively small average data set size. Outliers are to be expected due to the inconsistency of the data, and the sampling distribution is also likely to be sparse in some areas. These have a large negative effect on R-squared. Overall, the GPR models perform worse than the RF models in terms of pure predictive performance. This could be due to the actual inferiority of the GPR algorithm or the nature of the data, which are largely discontinuous as they come from different references. The latter presents a greater obstacle for the GPR models, which model with continuous distribution functions. The PCA pipeline improves the performance of the Gaussian process models but has mixed results for the random forest models, and the grid search pipeline generally performs well for both models.

Exploration leads to more broadly distributed training data and thus more representative samples. Therefore, it was also expected that this would lead to a better training base, which in turn would contribute to better generalization of the models.

The most accurate predictions at the end of SL can be achieved with the RF model. However, an R-square of 0.42 corresponds only to a rough estimate of the general trend in the formulations. While the grid search pipeline created solid baseline results for both models it could not deliver the same performance for the smaller training samples at the end of SL, which suggests overfitting. Vanilla models always perform best. Overall, the benchmarking of model performance does not provide any indication of the models' suitability for SL.

6. Discussion and conclusions

Compliance with the Paris climate agreement will not be possible without reducing emissions from cement production. The complex chemical reactions of their many components make it difficult to find useable alternatives. A classical experimental design approach reaches its limits, as thousands of experiments would be required to cover all possible combinations of starting materials. This is where SL comes into play: It shifts a large part of the task to ML predictions. Its acceleration potential lies in omitting unimportant experiments and finding materials with desired properties quicker through a targeted series of experiments. Compared to conventional data-based predictions, the emphasis is on data collection in the laboratory that is as parsimonious as it is effective.

The central hypothesis of this study was confirmed: The configuration of the SL-driven materials design, including the model selection, hyperparameter settings, and optimization strategy (as shown in Table 3), was found to have a significant impact on the expected performance. The best configuration resulted in a tenfold increase in performance compared to the least effective configuration (see Figs. 3 and 4).

The first research question was addressed by conducting an in-depth analysis of each parameter's influence, with the results summarized in Table 4. Based on these findings, an ideal strategy could be derived that aligns well with empirical results from SL benchmarking, as indicated in Table 5. Specifically, dimensionality reduction using PCA led to the highest algorithmic improvement overall, while for SOO, GPR with grid search produced the best performance. The most feasible scenario was found to be MOO between a prediction and an a-priori available information. Notably, incorporating a-priori information into the design is seldom mentioned in the literature, but it can significantly improve SL-optimization performance (see Figs. 3 and 4, right).

Regarding the second research question, the analysis showed that SL demonstrated significant acceleration compared to complete screening and the statistical RD baseline, as demonstrated by Figs. 3 and 4. In the optimal scenario, less than 5% of the formulations were sampled with SL to reach the target with high certainty, compared to 100% for complete screening and 25% for RD.

Finally, the analysis of the third research question showed that accurate predictions of material properties can be made in principle using the Random Forest (RF) algorithm, but only if a representative data

sample is available (as shown in Table 6). Surprisingly, the model prediction performance and SL optimization performance were not correlated, meaning that the best model performance did not necessarily lead to the best SL performance. While more advanced pipelines, which are aimed at improving model accuracy, can improve optimization performance in some scenarios, the pure predictive accuracy of the model does not seem to be a suitable indicator of overall usefulness. Here, simulated experiments are more appropriate because they directly measure the optimization performance regarding required development cycles.

These results challenge the usual emphasis on model accuracy in data-driven material development in civil engineering and show that valuable insights can be gained even with limited data sets and high input complexity. The exploration of different optimization scenarios and predictive accuracy led to relevant solutions, although AI currently cannot fully replace laboratory work. This promises immediate effectiveness in complex material development challenges.

The data collected also show that there does not appear to be a serious trade-off between environmental sustainability and the mechanical performance of materials. This suggests that designs can achieve high material quality while being environmentally sound. By incorporating a broader range of factors such as cost, resource availability, and other properties into materials development, this approach can revolutionize how complex building materials are researched and designed.

7. Limitations and future research potential

The key limitation of using ML in material science is its black-box nature. Despite the high confidence in the results produced by ML algorithms, researchers may need a deeper understanding of the relationships between the properties of a material and its performance. This requires careful configuration of the ML models. The success may be highly dependent on the quality and size of the training data. It is challenging to ensure that it is successful for the desired materials and applications. Nevertheless, this work has demonstrated that only a few iterations are necessary to achieve excellent optimization performance.

The data on which this study is based comes from numerous publications conducted in laboratories around the globe. This means that this data is likely to be more inconsistent than data from one lab. With more consistent data coming from one source, SL performance is expected to increase even further.

Furthermore, each reference contains a small number of samples on average leading to a fractured locally populated and sparse search space. In a real laboratory application, the search space would be populated more globally. This not only leads to a completely new set of research questions which cannot be addressed with the presented approach (such as the construction of the search space (Baird and Sparks, 2023) or multi-fidelity optimization (Palizhati et al., 2022)), but it may also put the conclusions that were drawn here into a novel perspective.

Given the constraints of data availability, our study focused on optimizing two target properties. However, the principles of SL and the statistical insights we gained are likely extendable to scenarios with more targets. Future research could explore these multi-target optimization scenarios, considering the regularization of target properties.

Future research should focus on developing data-driven approaches to building materials optimization, with the aim of automating the process and implementing the approach at larger scales. Although other materials science domains have seen success with automated platforms, there has been limited automation of experiments in the building materials sector. Consequently, research should focus on more high-throughput and/or automated experiments to optimize or discover new materials.

CRediT authorship contribution statement

Christoph Völker: Conceptualization, Methodology, Software, Formal analysis, Visualization, Supervision, Project administration, Funding acquisition, Writing – original draft. **Benjami Moreno Torres:** Data curation, Writing – original draft. **Tehseen Rug:** Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. **Rafia Firdous:** Writing, Data curation, Writing – original draft. **Ghezal Ahmad Jan Zia:** Writing – review & editing, Software, Formal analysis. **Stefan Lüders:** Writing – original draft, Software, Formal analysis. **Horacio Lisdero Scaffino:** Writing – original draft, Software. **Michael Höpler:** Software, Writing – review & editing. **Felix Böhmer:** Conceptualization, Project administration, Writing – review & editing. **Matthias Pfaff:** Funding acquisition, Writing – review & editing. **Dietmar Stephan:** Writing – review & editing. **Sabine Kruschwitz:** Writing – review & editing, Project administration, Funding acquisition.

Declaration of competing interest

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

Data availability

I have shared the data at the attach file step. The data and source code necessary to reproduce the results will be made available according to F.A.I.R. standards upon the acceptance of this work.

Green building materials: a new frontier in data-driven sustainable concrete design (Original data)DOI: 10.5281/zenodo.8081374 https://doi.org/10.5281/zenodo.8081374

Acknowledgements

We would like to thank Sterling Baird from the University of Utah for the many valuable discussions on the background of model optimization. We gratefully acknowledge support from the European Union's Horizon Europe research and innovation program under grant agreement No. 101056773 for funding this work under the Reincarnate project.

References

Abhilash, P., Sashidhar, C., Reddy, I., 2016. Strength properties of fly ash and GGBS based geo-polymer concrete. International Journal of ChemTech Research, Bd. 9 (3).

Adam, A.A., 2009. Thesis: Strength and Durability Properties of Alkali Activated Slag and Fly Ash-Based Geopolymer Concrete. School of Civil, Environmental and Chemical Engineering, Melbourne.

Addy, S., Koshla, N., 1986. „Computer Aided Design of Portland Cement Concrete Mixes. Transp Res Rec, pp. 63–69.

Alsalman, Ali, Assi, Lateef N., Kareem, Rahman S., Carter, Kealy, Ziehl, Paul, 2021. Energy and CO2 emission assessments of alkali-activated concrete and Ordinary Portland Cement concrete: a comparative analysis of different grades of concrete. Clean. Environ. Syst. 3.

Al-Tais, S.S., Annapurna, D., 2018. An experimental investigation on volcanic ash – based geopolymer concrete. International Journal of Engineering Technology Science and Research 5 (1).

Albitar, M., Visintin, P., Ali, M.M., Drechsler, M., 2015. Assessing behaviour of fresh and hardened geopolymer concrete mixed with class-F fly ash. Korean Society of Civil Engineering Journal of Civil Engineering, Bd. 19 (5), 1445–1455.

Aliabdo, A.A., Elmoaty, A.E.M.A., Salem, H.A., 2016a. Effect of cement addition, solution resting time and curing characteristics on fly ash based geopolymer concrete performance. Construct. Build. Mater. 123, 581–593.

Aliabdo, A.A., Elmoaty, A.E.M.A., Salem, H.A., 2016b. Effect of water addition, plasticizer and alkaline solution constitution on fly ash based geopolymer concrete performance. Construct. Build. Mater. 121, 694–703.

Aliabdo, A.A., Elmoaty, A.E.M.A., Emam, M.A., 2019. Factors affecting the mechanical properties of alkali activated ground granulated blast furnace slag concrete. Construct. Build. Mater. 197, 339–355.

Alsakka, F., Haddad, A., Ezzedine, F., Salami, G., Dabaghi, M., Hamzeh, F., 2023. „Generative design for more economical and environmentally sustainable reinforced concrete structures. Journal of Cleaner Production, Bd. 387, 135829 https://doi.org/10.1016/j.jclepro.2022.135829.

Andrew, R.M., 2018. Global CO2 emissions from cement production. Earth Syst. Sci. Data 195–217.

Aravindan, S., Jagadish, N., Guspher, A.P., 2015. „Experimental investigation of alkali-activated slag and flyash based geopolymer concrete. ARPN Journal of Engineering and Applied Sciences, Bd. 10 (Nr. 10), 4701–4705.

Baird, S., Sparks, T., 2023. „Spaces, Compactness Matters: Improving Bayesian Optimization Efficiency of Materials Formulations through Invariant Search," ChemRxiv (Pre-print), pp. 1–25. https://doi.org/10.26434/chemrxiv-2022-nz2w8-v3.

Barnard, R., 2014. Thesis: Mechanical Properties of Fly Ash/slag Based Geopolymer Concrete with the Addition of Macro Fibres. Stellenbosch University, Stellenbosch.

Bidwe, S.S., Hamane, A.A., 2015. Effect of different molarities of sodium hydroxide solution on the strength of geopolymer concrete. American Journal of Engineering Research (AJER) 4 (3), 139–145.

Bousquet, A., 2017. „lolopy [Online]. https://pypi.org/project/lolopy/. (Accessed 27 February 2023). Zugriff am.

Chaabene, W.B., Flah, M., Nehdi, M.L., 2020. „Machine Learning Prediction of Mechanical Properties of Concrete: Critical Review. Construction and Building Materials.

Chithambaram, S.J., Kumar, S., Prasad, M.M., Adak, D., 2018. Effect of parameters on the compressive strength of fly ash based geopolymer concrete. Struct. Concr. 19 (4), 1202–1209.

Comission, E., 2023a. „A European Green Deal Striving to Be the First Climate-Neutral Continent [Online]. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en. Zugriff am 28 02 2023].

Comission, European. „Coal regions in transition [Online]. https://energy.ec.europa.eu/topics/oil-gas-and-coal/eu-coal-regions/coal-regions-transition_en. (Accessed 24 March 2023). Zugriff am.

Fang, G., Ho, W.K., Tu, W., Zhang, M., 2018. Workability and mechanical properties of alkali-activated fly ash-slag concrete cured at ambient temperature. Construct. Build. Mater. 172, 476–487.

Firdous, R., Nikravan, M., Mancke, R., Vöge, M., Stephan, D., 2022. „Assessment of environmental, economic and technical performance of geopolymer concrete: a case study. J. Mater. Sci. 57, 18711–18725. https://doi.org/10.1007/s10853-022-07820-6.

Gallet, A., Rigby, S., Tallman, T.N., Kong, X., Hajirasouliha, I., Liew, A., D, L., Chen, L., Hauptmann, A., Smyl, D., 2022. „Structural engineering from an inverse problems perspective. Proc Math Phys Eng Sci., Bd. 478 (2257), 20210526 https://doi.org/10.1098/rspa.2021.0526.

Global Energy Monitor. Global coal plant tracker [Online]. https://globalenergymonitor.org/projects/global-coal-plant-tracker/. (Accessed 24 March 2013). Zugriff am.

Gökçe, H.S., Tuyan, M., Ramyar, K., Nehdi, M.L., 2020. „Development of eco-efficient fly ash–based alkali-activated and geopolymer composites with reduced alkaline activator dosage. Journal of Materials in Civil Engineering, Bd. 32 (2), 04019350, 10.1061/(ASCE)MT.1943-5533.0003017.

Golafshani, E.M., Arashpour, M., Kashani, A., 2021. „Green mix design of rubbercrete using machine learning-based ensemble model and constrained multi-objective optimization. J. Clean. Prod.

Haddad, R.H., Alshbuol, O., 2016. Production of geopolymer concrete using natural pozzolan: a parametric study. Construct. Build. Mater. 114, 699–707.

Hadi, M.N., Farhan, N.A., Sheikh, M.N., 2017. Design of geopolymer concrete with GGBFS at ambient curing condition using Taguchi method. Construct. Build. Mater. 140, 424–431.

Hardjito, D., Rangan, B.V., 2005. Development and Properties of Low-Calcium Fly Ash Based Geopolymer Concrete. Curtin University of Technology, Perth.

He, J., Jie, Y., Zhang, J., Yu, Y., Zhang, G., 2013. „Synthesis and characterization of red mud and rice husk ash-based geopolymer composites. Cement Concr. Compos. 37, 108–118. https://doi.org/10.1016/j.cemconcomp.2012.11.010.

Hongen, Z., Feng, J., Qingyuan, W., Ling, T., Xiaoshuang, S., 2017. Influence of cement on properties of fly-ash-based concrete. ACI Mater. J. 114 (5), 745–753.

Honggen, Z., Ling, T., Xiaoshuang, S., Qingyuan, W., 2017. „Influence of curing condition on compressive strength of low-calcium fly ash-based geopolymer concrete. Journal of Residuals Science and Technology, Bd. 4 (Nr. 1), 79–83.

Jeyasehar, C.A., Saravanan, G., Salahuddin, M., Thirugnanasambandam, S., 2013. Development of fly ash based geopolymer precast concrete elements. Asian Journal of Civil Engineering (BHRC) 14 (4), 605–615.

K, N.V., Babu, D.V., 2018. Assessing the performance of molarity and alkaline activator ratio on engineering properties of self_compacting alkaline activated concrete at ambient temperature. J. Build. Eng. 20, 137–155.

Kim, Y., Kim, E., Antono, E., Meredig, B., Ling, J., 2020. Machine-learned metrics for predicting the likelihood of success in materials discovery. npj Comput Mater, Bd. 6 (131), 6020. https://doi.org/10.1038/s41524-020-00401-8.

Kotlar, A.M., Iversen, B.V., de Jong van Lier, Q., 2019. „Evaluation of parametric and nonparametric machine-learning techniques for prediction of saturated and near-saturated hydraulic conductivity. Vadose Zone Journal, Bd. 18, 1–13.

Kumar, S.S., Vasugi, J., Ambily, P., Bharatkumar, B.H., 2013. Development and Determination of Mechanical properties of fly ash and slag blended geopolymer concrete. Int. J. Sci. Eng. Res. 4 (8).

Li, Z., Yoon, J., Zhang, R., Rajabipour, F., Srubar III, W.V., Dabo, I., Radliska, A., 2022. Machine learning in concrete science: applications, challenges, and best practices. npj Comput. Mater. 1–17.

Liang, Q., Gongora, A.E., Ren, Z., Tiihonen, A., Liu, Z., Sun, S., Deneault, J.R., Bash, D., Mekki-Berrada, F., Khan, S.A., Hippalgaonkar, K., Maruyama, B., Brown, K.A., Fisher III, J., Buonassisi, T., 2021. „Benchmarking the performance of Bayesian optimization across multiple experimental materials science domains. npj Computational Materials, Bd. 7 (1) https://doi.org/10.1038/s41524-021-00656-9, 1: 10.

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, pp. 207–217.

Lookman, T., Balachandran, P.V., Xue, D., Yuan, R., 2019. „Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. npj Computational Materials, Bd. 5, 1–17.

Manickavasagam, R., Mohankumar, G., 2017. Studyon high calcium flyash based geopolymer concrete. American Journal of Engineering Research (AJER) 6 (1), 86–90.

Mathieson, J.G., Somerville, M.A., Deev, A., Jahanshahi, S., 2015. Utilization of biomass as an alternative fuel in ironmaking. Iron Ore, Bd. 42, 581–613. https://doi.org/10.1016/B978-1-78242-156-6.00019-8.

Memon, F.A., Nuruddin, M.F., Demie, S., Shafiq, N., 2011. Effect of curing conditions on strength of fly ash-based self-compacting geopolymer concrete. Int. J. Civ. Environ. Eng. 5 (8), 342–345.

Memon, F.A., Nuruddin, M.F., Shafiq, N., 2013. Effect of silica fume on the fresh and hardened properties of fly ash-based self-compacting geopolymer concrete. Int. J. Miner. Metall. Mater. 20 (2), 205–213.

Montoya, J.H., Winther, K.T., Flores, R.A., Bligaard, T., Hummelsh, J.S., Aykol, M., 2020. „Autonomous intelligent agents for accelerated materials discovery. Chemical science, Bd. 11.

Muthadhi, A., Vanjinathan, J., Durai, D., 2016. Experimental investigations on geo polymer concrete based on class C fly ash. Indian Journal of Science and Technology 9 (5), 1–5.

Naderpour, H., Rafiean, A.H., Fakharian, P., 2018. „Compressive strength prediction of environmentally friendly concrete using artificial neural networks. J. Build. Eng. 213–219.

Nath, P., Sarker, P.K., 2014. Effect of GGBFS on setting, workability and early strength properties of fly ash geopolymer concrete cured in ambient condition. Construct. Build. Mater. 66, 163–171.

Nath, P., Sarker, P.K., 2015. Use of OPC to improve setting and early strength properties of low calcium fly ash geopolymer concrete cured at room temperature. Cement Concr. Compos. 55, 205–214.

Nath, P., Sarker, P.K., 2017a. Flexural strength and elastic modulus of ambient-cured blended low-calcium fly ash geopolymer concrete. Construct. Build. Mater. 130, 22–31.

Nath, P., Sarker, P.K., 2017b. Fracture properties of GGBFS-blended fly ash geopolymer concrete cured in ambient temperature. Mater. Struct. 50 (32).

Noushini, A., Castel, A., 2016. The effect of heat-curing on transport properties of low-calcium fly ash-based geopolymer concrete. Construct. Build. Mater. 112, 464–477.

Olivia, M., Nikraz, H., Sarker, P., 2011. Improvements in the strength and water penetrability of low calcium fly ash based geopolymer concrete. In: ARPN Journal of Engineering and Applied Sciences, Ipswich.

Omar, O.M., Elhameed, G.D.A., Heniegal, A.M., Mohamadien, H.A., 2015. Effect of crashed stone on properties of fly ash based-geopolymer concrete with local alkaline activator in Egypt. Int. J. Civ. Environ. Eng. 9 (10), 1357–1366.

Palizhati, A., Torrisi, S.B., Aykol, M., Suram, S.K., Hummelshøj, J.S., Montoya, J.H., 2022. Agents for sequential learning using multiple-fidelity data. Scientific Reports, Bd. 12 (1), 2045–2322. https://doi.org/10.1038/s41598-022-08413-8.

Pan, Z., Sanjayan, J.G., Rangan, B.V., 2011. Fracture properties of geopolymer paste and concrete. Mag. Concr. Res. 63 (10), 763–771.

Parthiban, K., Mohan, S.R., 2014. Effect of sodium hydroxide concentration and alkaline ratio on the compressive strength of slag based geopolymer concrete. Int. J. ChemTech Res. 6 (4), 2446–2450.

Parthiban, K., Vaithianathan, S., 2015. Effect of kaolin content and alkaline concentration on the strength development of geopolymer concrete. Int. J. ChemTech Res. 8 (4), 1730–1734.

Pavithra, P., Reddy, M.S., Dinakar, P., Rao, B.H., Satpathy, B., Mohanty, A., 2016. A mix design procedure for geopolymer concrete with fly ash. J. Clean. Prod. 133, 117–125.

Provis, J.L., van Deventer, J., 2014. Alkali activated materials, state-of-the-art report. In: RILEM TC 224-AAM. Springer, Dordrecht, Heidelberg, New York, London.

Provis, J.L., Palomo, A., Shi, C., 2015. „Advances in understanding alkali-activated materials. Cement Concr. Res. 110–125.

Rafeet, A., Vinai, R., Soutsos, M., Sha, W., 2017. „Guidelines for mix proportioning of fly ash/GGBS based alkali activated concretes. Construction and Building Materials, Bd. 147, 130–142.

Rai, B., Roy, L.B., Rajjak, M., 2018. A statistical investigation of different parameters influencing compressive strength of fly ash induced geopolymer concrete. Struct. Concr. 19 (5), 1268–1279.

Rao, G.M., Rao, T.D.G., 2018. A quantitative method of approach in designing the mix proportions of fly ash and GGBS-based geopolymer concrete. Aust. J. Civ. Eng. 16 (1), 23–63.

Raschka, S., 2018. „MLxtend: providing machine learning and data science utilities and extensions to Python's scientific computing stack. The Journal of Open Source Software, Bd. 3.

Reddy, M.S., Dinakar, P., Rao, B.H., 2018. Mix design development of fly ash and ground granulated blast furnace slag based geopolymer concrete. J. Build. Eng. 20, 712–722.

Reyes, K., Powell, W.B., 2020. Optimal Learning for Sequential Decisions in Laboratory Experimentation arXiv preprint arXiv:2004.05417.

Rohr, B., Stein, H.S., Guevarra, D., Wang, Y., Haber, J.A., Aykol, M., Suram, S.K., Gregoire, J.M., 2020. „Benchmarking the acceleration of materials discovery by sequential learning. Chem. Sci., Bd. 11 (10), 696–2706. https://doi.org/10.1039/C9SC05999G.

Sarker, P.K., Haque, R., Ramgolam, K.V., 2013. Fracture behaviour of heat cured fly ash based geopolymer concrete. Mater. Des. 44, 580–586.
Scikit learn, „scikit learn [Online]. https://scikit-learn.org/stable/modules/generated/sklearn.gaussian_process.GaussianProcessRegressor.html. (Accessed 15 December 2022). Zugriff am.
Sergis, V., Ouellet-Plamondon, C.M., 2022. Automating mix design for 3D concrete printing using optimization methods. Digital Discovery 645–657.
Shehab, H.K., Eisa, A.S., Wahba, A.M., 2016. Mechanical properties of fly ash based geopolymer concrete with full and partial cement replacement. Construct. Build. Mater. 126, 560–565.
Shinde, B.H., Kadam, K.N., 2016. Strength properties of fly ash based geopolymer concrete with sea sand. American Journal of Engineering Research (AJER) 5 (7), 129–132.
Shobeiri, V., Bennett, B., Xie, T., Visintin, P., 2022. A generic framework for augmented concrete mix design: optimisation of geopolymer concrete considering environmental, financial and mechanical properties. Journal of Cleaner Production, Bd 369, 133382. https://doi.org/10.1016/j.jclepro.2022.133382.
Song, Y., Wang, X., Li, H., He, Y., Zhang, Z., Huang, J., 2022. „Mixture optimization of cementitious materials using machine learning and metaheuristic algorithms: state of the art and future prospects. Materials, Bd. 15, 7830. https://doi.org/10.3390/ma15217830.
Steel, R.G.D., Torrie, J.H., 1960. Principles and procedures of statistics. With special Reference to the Biological Sciences 4 (1), 207–208. https://doi.org/10.1002/bimj.19620040313.
Sun, Q., Zhu, H., Li, H., Zhu, H., Gao, M., 2018. Application of response surface methodology in the optimization of fly ash geopolymer concrete. Revista Romana de Materiale/Romanian J. Mater 48 (1), 45–52.
Topark-Ngarm, P., Chindaprasirt, P., Sata, V., 2014. „Setting time, strength, and bond of high-calcium fly ash geopolymer concrete. Journal of Materials in Civil Engineering, Bd. 27 (7).
Vinai, R., Rafeet, A., Soutsos, M., 2016. „The role of water content and paste proportion on physico-mechanical properties of alkali activated fly ash–ggbs concrete. Journal of Sustainable Metall, Bd. 2, 51–61.
Völker, C., Völker, A., 2022. Sequential Learning App for Materials Discovery ("SLAMD") [Online]. https://github.com/BAMresearch/SequentialLearningApp.
Völker, C., Firdous, R., Stephan, D., Kruschwitz, S., 2021. „Sequential learning to accelerate discovery of alkali-activated binders. J. Mater. Sci. 15859–15881.
Völker, C., Kruschwitz, S., Moreno Torres, B., Firdous, R., Zia, G.A., Stephan, D., 2022. „Accelerating the search for alkali-activated cements with sequential learning. In: Fib International Congress. Norway, Oslo.
Von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Walczak, M., Pfrommer, J., Pick, A., 2021. „Informed Machine Learning-A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems. IEEE Transactions on Knowledge and Data Engineering.
Vora, P.R., Dave, U.V., 2013. Parametric studies on compressive strength of geopolymer concrete. In: Chemical, Civil and Mechanical Engineering Tracks of 3rd Nirma University International Conference on Engineering. NUiCONE-2012).
Wager, S., Hastie, T., Efron, B., 2014. Confidence intervals for random forests: the jackknife and the infinitesimal jackknife. J. Mach. Learn. Res. 15, 1625–1651.
Wardhono, A., 2015. Thesis: the Durability of Fly Ash Geopolymer and Alkali-Activated Slag Concretes. RMIT University, Melbourne.
Xie, T., Visintin, P., Zhao, X., 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. Construct. Build. Mater., 119380 https://doi.org/10.1016/j.conbuildmat.2020.119380, 256 17. April 2020.
Yang, K.-H., Song, J.-K., Song, K.-I., 2017. „CO2 reduction assessment of alkali-activated concrete based on Korean life-cycle inventory database. In: Handbook of Low Carbon Concrete. Elsevier, pp. 139–157.
Ye, A., Wang, Z., 2023. Modern Deep Learning for Tabular Data - Novel Approaches to Common Modeling Problems. USAQ: Apress, Berkeley, CA.
Zhang, J., Huang, Y., Wang, Y., Ma, G., 2020. „Multi-objective optimization of concrete mixture proportions using machine learning and metaheuristic algorithms. Construct. Build. Mater. 253, 119208 https://doi.org/10.1016/j.conbuildmat.2020.119208.