Construction and Building Materials 244 (2020) 118271

Contents lists available at ScienceDirect

Construction and Building Materials

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

An ensemble machine learning approach for prediction and optimization of modulus of elasticity of recycled aggregate concrete

Taihao Han [a], Ashfia Siddique [b], Kamal Khayat [b], Jie Huang [c], Aditya Kumar [d,*]

[a] Department of Materials Science and Engineering, Missouri University of Science and Technology, Rolla, MO, USA
[b] Department of Civil, Architectural, and Environmental Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA
[c] Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA
[d] Department of Materials Science and Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA

HIGHLIGHTS

- This study presents the first application of an ensemble machine learning (ML) model to predict the modulus of elasticity (MOE) of recycled aggregate concrete.
- The ensemble ML model – comprising of random forests (RF) and support vector machine (SVM) – produces accurate predictions of concretes' MOE (RMSE of ≈3.0 GPa).
- Prediction performance of the ensemble ML model is consistently superior than several standalone ML models.
- The ensemble ML model is able to develop optimal mixture designs for RCA concretes that satisfy imposed target MOE.

ARTICLE INFO

Article history:
Received 15 July 2019
Received in revised form 23 October 2019
Accepted 22 January 2020

Keywords:
Recycled concrete aggregate (RCA)
Modulus of elasticity (MOE)
Ensemble machine learning
Random forests
And voting

ABSTRACT

This paper presents an ensemble machine learning (ML) model for prediction of modulus of elasticity (MOE) of concrete formulated using recycled concrete aggregate (RCA), in relation to features of its mixture design (e.g., physiochemical characteristics of RCA). The ensemble ML model's prediction performance was compared with five commonly-used ML models. It is shown that the ensemble ML model unfailingly produces more accurate predictions compared to standalone models. To demonstrate the ability of the ensemble ML model to go beyond MOE predictions, the model was used to develop optimal mixture designs for RCA concretes that satisfy imposed target MOE.

1. Introduction

Solid waste produced from demolition of construction infrastructure – that is, Construction and Demolition Waste (CDW) – continues to grow in volume, at a rapid pace, in not just the United States (U.S.) but across the world [1]. According to the U.S. Environmental and Protection Agency (EPA), 535 million tons of CDW were generated in the U.S. in 2014, a 65% increase from 2003 [2,3]. In the past, when applications for CDW had not been conceived, the waste material was typically discarded and disposed in landfills, thus resulting in consumption of valuable land resources that could otherwise have been used for beneficial purposes (e.g., afforestation and new construction). Notwithstanding, in the past decade or so, there has been burgeoning interest in recycling CDW (i.e., consolidating and grinding into smaller particulates) and using the recycled product in various applications (e.g., as backfilling material and aggregate for concrete) [4], essentially to address concerns pertaining to progressive reduction in landfilling capacity as well as to alleviate the exorbitant cost associated with transportation of CDW to landfilling sites. The EPA estimates that roughly 70% of CDW can be classified as concrete [1,2]. During recycling of CDW, the waste concrete is crushed and processed into three different classes of aggregates: coarse recycled concrete aggregate (coarse RCA, particulate size >4.75 mm); fine RCA (particulate size between 0.075 and 4.75 mm); and recycled concrete fines (RCF, particulate size <0.075 mm) [5,6]. Of these three classes

* Corresponding author at: Department of Materials Science and Engineering, Missouri University of Science and Technology, B49 McNutt Hall, 1400 N Bishop, Rolla, MO 65409, USA.

E-mail addresses: thy3b@mst.edu (T. Han), aspc8@mst.edu (A. Siddique), khayat@mst.edu (K. Khayat), jieh@mst.edu (J. Huang), kumarad@mst.edu (A. Kumar).

https://doi.org/10.1016/j.conbuildmat.2020.118271
0950-0618/

of waste concrete, coarse RCA is the major component, embodying 70%–85%$_{mass}$ of the material. This study is focused on the influence of coarse RCA (subsequently referred to as RCA) on properties of concrete.

Glushge, in his pioneering work [7], posited that RCA could be used to replace – either partially or exhaustively – the natural aggregate that is used in concrete. The original idea was based on the premise that concrete, given its enormous scale of production and use (≈46 billion tons per year [8–10]), would serve not just as an ideal but also a proportionately and adequately large repository for RCA (which, like concrete, is produced at a large scale, and has widespread and abundant availability). Furthermore, replacement of natural aggregate – sourced from natural resources – with RCA would assuage, at least to some extent, the alarming rate at which such resources are currently being depleted (primarily, to meet the enormous demand of concrete's production and use). Lastly, if locally-produced RCA is used in concrete, costs associated with transportation and disposal of RCA would be significantly reduced. This would – as an added benefit – free up land resources, making them available for beneficial purposes as opposed to being used exclusively for disposal of CDW.

While RCA, based on its nomenclature, is classified as *aggregate*, its physiochemical characteristics are disparate from those of natural aggregate; therefore, its use in concrete is more complex than the mere exercise of substituting natural aggregate with RCA. The primary reason for such disparity in properties of RCA with respect to those of natural aggregates is that the particulates of the former material are comprised of natural coarse aggregates, the surfaces of which are either partially or fully covered by a layer of mortar (i.e., hydrated cement paste intermixed with sand) [11–18]. The encompassing surface layer of mortar is porous and, consequently, highly permeable; as a result, the water absorption capacity of RCA is generally higher than that of natural aggregate [5]. Furthermore, on account of larger surface area of RCA particulates (due to surface roughness of the mortar layer), concrete provisioned with RCA (subsequently referred to as RCA concrete) has volumetrically larger interfacial transition zones (i.e., ITZs: water-filled zones around the RCA particulates) and, thus, higher water demand compared to an equivalent conventional concrete (formulated using natural aggregate) to produce a specific level of workability. Depending on the strength of bonds at the aggregate-mortar interface, various mechanical properties (e.g., stiffness and restraining capacity) of RCA particulates could be rendered slightly to significantly inferior in comparison to those of equivalent-volume natural aggregate particulates [17]. Owing to the intrinsically poorer mechanical properties of RCA vis-à-vis those of natural aggregate, mechanical properties of RCA concrete are often different – in general, inferior – compared to its counterpart formulated using natural aggregate [11–17]. If the aforementioned aggregate-mortar bonds at the topographical sites of RCA particulates are weak or if the RCA has high water absorption capacity (leading to large ITZs), the overall rigidity of RCA concrete is diminished, which, in turn, reduces the concrete's modulus of elasticity (MOE) thereby making it more susceptible to deformation [7,12,17,19]. The presence of microcracks in RCA particulates, or in the mortar covering RCA particulates' surfaces, could result in even further reduction in RCA concrete's MOE. At critically low MOE, RCA concrete (or any other concrete for that matter) becomes highly susceptible to deformation, cracking, and failure when subjected to environmental and/or mechanical loads – even those of small magnitudes and even if they are applied only for short durations [5].

Based on the above discussion, it is clear that partial or complete substitution of natural aggregate with RCA could have profound effect on the performance of concrete [20]. Among the various metrics of concrete's performance, MOE is deemed a significant compliance criterion, and often used as one of the principal specification criteria to determine constructability of a given structure [5,21]. In the specific case of RCA concrete, assessment of MOE gains even more importance because of the overwhelming evidence in literature [11–17] suggesting that with increasing substitution of natural aggregate with RCA there is proportional decline in MOE of concrete. While important, determination of MOE of concretes – even conventional ones, leave alone those formulated using RCA – using semi-empirical methods (e.g., a mathematical function listed in [22]) is unreliable [5]. This is because the foundational mechanisms, which are at the origin of concrete's MOE and its dependency on other features of concrete (e.g., compressive strength, as suggested in [22]), are not well understood and, therefore, not accurately accounted for in the aforesaid semi-empirical formalisms. Several studies have employed unconstrained/unsupervised statistical approaches [7,12,23–25] in attempts to develop simple mathematical functions that can predict MOE of RCA concrete in relation to other properties of the material that can be readily measured (e.g., unit weight and compressive strength). However, due to the staggeringly large compositional degrees of freedom in RCA concretes – arising from various permutations and combinations of mixture design variables and physiochemical characteristics (e.g., density and maximum particulate size) of RCA – coupled with the inherent nonlinear relationships between mixture design variables and concrete's properties, such statistically-derived functions are not generically applicable, often failing to make accurate predictions of MOE of RCA concretes that are compositionally different from those used to develop the functions [23,25].

In light of the abovementioned deficiencies of unsupervised and unconstrained statistical models, several researchers have focused on developing and employing machine learning (ML) frameworks – both supervised and unsupervised – for prediction, and in some cases optimization, of concrete properties, including MOE. The broad appeal of ML models is befitting because such models – provided that they are trained rigorously using high-quality datasets – are able to not only reveal the elementary (yet hidden) semi-empirical rules which dictate the rudimentary linkage between concrete properties and mixture design, but also perform predictions in previously untrained data-domains [26,27]. Majority of past ML-based studies have placed emphasis on prediction of compressive strength of conventional concretes [26–40] using their physiochemical attributed [e.g., cement content; water content; and mass/volume of admixture and/or mineral additive] and age as inputs; only a handful of articles have been focused on prediction of mechanical properties of RCA concrete. Duan et al. [41] applied a nonlinear, regression-based ML model, that is, artificial neural network (ANN), to predict the 28-day MOE and compressive strength of RCA concrete. Gholampour et al. [42] examined the applicability of various regression-based ML models – multivariate adaptive regression splines (MARS); M5 model tree (M5Tree); and least squares support vector regression (LSSVR) – to predict mechanical properties of RCA concrete, including age-dependent MOE. Deshpande et al. [43] applied ANN to predict compressive strength of RCA concrete, which, when combined with semi-empirical equations, can potentially be used to estimate MOE. Behnood et al. [5] used the M5P model tree algorithm – a relatively new decision tree ML model [44] – to predict the MOE of RCA concrete. Deng et al. [45] employed a convolutional ANN based deep learning model to predict compressive strength of RCA concrete. Sadati et al. [18] used multilayer perceptron artificial neural network (MLP-ANN) – a subset of artificial neural network, consisting of multiple hidden layers of neurons – for prediction and optimization of MOE of RCA concrete. In all of the studies cited here, it has been shown that nonlinear, regression-based ML models, upon adequate training and rigorous testing, are able to predict mechanical properties of RCA concrete with reasonable accuracy (i.e., Pearson correlation coefficient, R, up to 0.91). Nonetheless, it is important to point out that the two most-used ML models in prior studies – that is, ANN and M5P models – are often falter at accurately predicting outcomes in data-domains that feature complex (e.g., greatly nonlinear and/or non-monotonic) input(s)-output(s) relationship [46–49]. This is because ANN models – as well as its derivatives such as MLP-ANN – are premised on local optimization and search algorithms (e.g., the back-propagation mechanism that is used in several neural network-based ML models for optimization of activation functions' parameters) that are highly vulnerable in terms of getting confined in (or around) a local minima as opposed to converging to the global minimum [46–49]. Owing to this issue, ANN models often produce disparate – even inferior – predictions for the same set of inputs when they are re-trained (e.g., using a larger or a different database). The other popular model, that is, the M5P model, uses logical-operators to split the data, and then employs multivariate linear functions to develop input–output correlations [50,51]. Because of its inherent reliance on linear – as opposed to nonlinear functions – the M5P model's prediction performance in complex data-domains (e.g., sinusoidal or logistic or dampening functions) could be poor [49]. In recent studies [27,49], it has been shown that the random forests (RF) model – a ML model, based on modification of the bootstrap aggregation decision tree algorithm – outperforms other standalone ML models, in terms of accuracy of prediction of compressive strength of concrete. These studies have attributed the RF model's superior prediction performance to its unique capacity to manage discrete as well as continuous variables over both monotonic and non-monotonic data-domains [52], while reducing variance among different subsets of the training dataset. In spite of the merits of the RF model, based on the authors' extensive literature review, it was found that the RF model has yet never been employed to predict MOE, or other properties, of RCA concrete.

The study presents the first application of an ensemble ML model – developed by combining the RF model with the support vector machine (SVM) model – to predict MOE of RCA concrete in relation to its mixture design and physiochemical attributes (e.g., particle size and water absorption capacity) of the RCA. The accuracy of predictions of the ensemble ML model is benchmarked against several standalone ML models [i.e., multilayer perceptron ANN (MLP-ANN), Gaussian Process Regression (GPR), linear regression (LR), SVM, and RF] that have been used in prior studies to predict properties of heterogeneous materials. For assessment of prediction performance of ML models, a real-world database, comprised of >500 unique mixture designs (and RCA characteristics) of RCA concretes and their corresponding (experimentally measured) 28-day MOE, are used for training and testing of the models. Five distinct statistical variables, a singular composite performance index (CPI) derived from the aforesaid variables, are used to quantitatively assess the ability of each ML model to predict in previously untrained compositional domains. The input–output correlation developed by the ensemble ML model – based on its training from the real-world database – is ultimately used to develop optimal mixture designs of RCA concretes that satisfy different imposed target MOEs (i.e., 30–50 GPa).

The paper is structured as follows. Section 2 presents brief description of the individual and ensemble ML models. In Section 3, databases used for training and testing of the ML models are described. Section 4 reports the comparison of prediction performances of various ML models as well as the results obtained from optimizations. Section 5 summarizes the main conclusions drawn from this study.

2. Overview of machine learning (ML) models

In this section, a succinct overview of five standalone and the ensemble ML model, implemented in this study, is presented. Further details pertaining to all six ML models are provided in **Research Data (Mendeley)**.

Support vector machine (SVM), a commonly used ML model for both classification and regression purposes, approximates the correlation – either in the form of multivariate linear or nonlinear functions – between inputs and output of a dataset. This is accomplished by employing an optimization scheme – as opposed to a regression approach – geared towards minimizing an objective cost function (i.e., ε-insensitive loss function), or simply put, to transform input data into a higher-dimensional structure such that data with similar characteristics are sequestered from dissimilar ones [53]. Artificial neural network – abbreviated as ANN – consists of multiple neurons arranged in hierarchical layers. Each of the neurons serves as a computational element and is responsible for processing information relayed from the previous layer of neurons (using sigmoidal or logistic-transfer activation functions) and transmitting the processed information to the next layer of neurons [54]. The structure of ANN resembles the network of interconnected neurons within the human brain – wherein information is processed (and simplified) and transmitted from one layer to another in a hierarchical manner. Multilayer perceptron artificial neural network (MLP-ANN) is a subset of ANN with multiple layers of neurons, and, therefore, strong self-learning capabilities [55]. Linear regression (LR) is a simple ML technique that uses piecewise linear functions – driven by independent predictors – to predict a numerical target based on a set of independent inputs [56]. Gaussian process regression (GPR) is a regression algorithm based on the Bayes' theorem, which identifies the most probable outcome (or hypothesis) on the basis of prior knowledge acquired from the training database [57]. The GPR model employs a stochastic process to collect random variables, any finite number of which have a joint Gaussian distribution [58]. The variables – which represent prior knowledge – are used to estimate the probability of a given outcome, and compare it against the probabilities of all possible outcomes. Through such comparisons, the prior knowledge is updated iteratively throughout the training process; ultimately, the outcome with the highest probability is selected as the final prediction. Random forests (RF) is a modification of [decision tree algorithm + bootstrap aggregation], premised on amalgamation of *bagging* and *adaptive nearest neighbors* to achieve logic-based inference of input–output correlations in a dataset. RF employs two-stage randomization to grow a large number of uncorrelated "deep" trees, all without any pruning or smoothening (unlike conventional decision trees-based models, which do require pruning during the training process) [59,60].

Voting is an ensemble method that uses a voting-based approach to combine predictions from two or more ML models into a singular prediction. In weighted voting [61], the contribution (measured in the form of vote) of each ML model is ascertained based on its prediction accuracy over the training dataset. More specifically, ML models with higher prediction accuracy are assigned superior weights than those with lower prediction accuracy. The final prediction is obtained by summing up all the votes – along with their weights – and selecting the ML model with superior aggregate vote compared to the others. In this study, the weighted voting method based on the combination of RF and SVM models was chosen.

The authors would like to point out that all standalone ML models described above comprise of hyper-parameters that needs to be adjusted by the user to improve their prediction performance. In some models, user intervention is also required to select optimal functions (e.g., type of kernel function used for transformation of dimensionality of data in SVM; and type of activation function used for activation of neurons in MLP-ANN). In this study, for the selection of optimal functions and hyper-parameters for each ML model, 10-fold cross-validation (CV) method [26,62,63] was used. In short, the 10-fold CV method randomly splits the training database into 10 equisized folds. The ML model – the hyper-parameters/functions of which need to be optimized – is trained using data-records from 9 folds, and subsequently blind-tested against data-records in the 10th fold. This process is iteratively repeated 9 times – each time using a unique combination of folds for training of the ML model and its blind-testing. During each iteration, the relevant parameters and functions of the ML models are fine-tuned such that prediction errors [measured in terms of root mean squared error (RMSE)] are progressively minimized. Functions and parameters' values after the last iteration are selected as "optimum," and used – without further changes – for subsequent testing of the ML model against a blind test dataset.

3. Development of database and assessment of performance of ML models

3.1. Development of database

Experimental data-records, collated from literature and original experiments, were used to train the ML models (described in Section 2), assess their prediction performance in previously untrained data-domains, and to conduct optimizations. The database comprised of 526 unique data-records; 483 data-records were mined (by Sadati et al. [18]) from published articles (which are enumerated in [18]) and another 43 were obtained from original experiments. In the context of the said original experiments [18], two different types of cementitious binders were used for formulation of RCA concretes: plain and blended (binary/ternary) binders. In plain binders, Type I/II OPC, without any substitution, was used as the binding cementitious material. In binary binders, 25% of OPC was replaced with class C fly ash, and in ternary binders, 50% of OPC was replaced with class C fly ash (35%) and blast furnace slag (15%). The amount of water was varied so as to maintain a water to cementitious materials mass ratio (w/cm) of 0.37 (low), or 0.40 (intermediate), or 0.45 (high). RCA (coarse recycled concrete aggregates) were obtained from six different sources – five recycling centers and one manufactured in our own laboratory – each with different physiochemical characteristics (i.e., density, maximum particulate size, and water absorption capacity). RCA was used to replace 0–100%$_{\text{mass}}$ of the natural coarse aggregate (i.e., mixture of crushed dolomite and limestone) in concrete. Siliceous river sand was used as fine aggregate. Slump and air content were adjusted according to ASTM C 143 [64] and ASTM 231 [65], using commercially-available water reducing admixture and air-entraining agent, respectively. MOE of the concrete specimens were evaluated using ASTM C 469 [66]; towards this, cylindrical concrete specimens, with dimensions of 100×200 mm (4×8 in.), were cast and cured in saturated lime solution at 21 ± 2 °C for 28 days until the test. To ensure consistency between data mined from literature and those obtained from experiments [18], emphasis was given to ensure that: (i) components of RCA concrete [e.g., types of cement, supplementary cementitious materials (SCMs), and RCA]; (ii) curing conditions of concrete; and (iii) protocols used for evaluation of concrete' MOE were broadly similar.

By consolidating both datasets (i.e., obtained from literature and experiments), a singular database was devised; within the database, each data-record had 13 inputs and 1 output. The 13 inputs parameters included: type of binder ("0" for plain binder and "1" for binary/ternary binder); contents (in kg·m^{-3}) of cement, SCMs (i.e., class C fly ash or a combination of class C fly ash and blast furnace slag), natural coarse aggregate, RCA, fine aggregate, and water; and density (in kg·m^{-3}), water absorption capacity (in %) and (nominal) maximum aggregate size (in mm) of the natural coarse aggregate and RCA. The output parameter included the 28-day MOE (in GPa) of all concrete mixtures. Statistical parameters pertaining to the database are shown in Table 1.

It was previously stated in Section 1 that in RCA concrete the relationships between input variables and output are complex (e.g., highly nonlinear and non-monotonic), thus necessitating the use of ML models to reveal them in order to make predictions and/or conduct optimizations. To corroborate this argument, Figs. 1 and 2 are shown, wherein it is evident from the two- (2D) and three-dimensional (3D) plots that the correlation between MOE and input variables is indeed complex. Such complexity, however, is foreseeable because each input variable – pertaining to concrete mixture design or RCA's characteristics – has a distinct, and profound, effect on the concrete's MOE. When multiple input variables are concurrently changed – for example, in Figs. 1 and 2 – deconvolution of the various effects (exerted by the different input variables) from 2D/3D plots or simple (semi-)empirical relationships is not feasible.

3.2. Assessment of performance of ML models

For the purposes of training, and to assess the prediction performance, of ML models, the experimental database (described in Section 3) was randomly apportioned into two subsets: one for training and the other for testing. 75% of data-records of the parent database (i.e., training set) were used to rigorously train (i.e., to select and optimize functions and hyper-parameters) the ML models; the remaining 25% of the data-records (i.e., testing set) were utilized for assessment of prediction performance of the models. Various past studies [26,27,30,49] have also used such split of 75%–25% in the parent database for training and testing of ML models. It is clarified that while splitting of the parent database was done in a randomized manner, care was taken to ensure that the training dataset was archetypal – albeit a shortened version – of the parent database. To this end, the training set was formulated in manner that each input variable spanned over a wide range between (and excluding) its minimum and maximum values in the parent database.

To quantitatively measure the accuracy of prediction produced by the ML models (against the test set), 5 distinct statistical measures were used; these parameters were extracted through comparisons of the models' predictions against actual measurements. These parameters include: coefficient of determination (R^2); root mean squared error (RMSE); Person correlation coefficient (R); mean absolute percentage error (MAPE); and mean absolute error (MAE). Equations used to calculate the aforesaid parameters can be found in [49].

$$CPI = \frac{1}{N}\sum_{j=1}^{j=N}\frac{P_j - P_{min,j}}{P_{max,j} - P_{min,j}} \tag{1}$$

The five statistical parameters, described above, were integrated into the composite performance index (CPI, see Eq. (1)) [26,49,67], to obtain a singular, unified measure of prediction performance of each ML model. In Eq. (1): N (=5) is the total number of statistical parameters used to measure performance of the models; P_j is magnitude of the j^{th} statistical parameter; and $P_{j,\,min}$ and $P_{j,max}$ are the minimum (i.e., worst) and maximum (i.e., best) values of the j^{th} statistical parameter. With Eq. (1) formulated in the manner described here, CPI of any given ML model can vary between 0 and 1. The best ML model would acquire a CPI value of 0 (or the lowest

Table 1
A summary of statistical parameters pertaining to each of the 14 attributes (13 input and 1 output) of the database. The database consists of 526 unique data-records (based on [18]).

Attribute	Unit	Min.	Max.	Mean	Std. Dev.
Binder type	Unitless	0	1	–	–
Cement (OPC) content	kg·m $^{-3}$	150.00	597.00	338.68	77.21
SCM (fly ash and/or slag) content	kg·m $^{-3}$	0.00	225.09	32.32	57.82
Natural aggregate (coarse) content	kg·m $^{-3}$	0.00	1950.00	563.09	434.25
RCA (coarse) content	kg·m $^{-3}$	0.00	1800.00	495.38	423.50
Fine aggregate content	kg·m $^{-3}$	465.00	1301.10	730.69	121.87
Natural agg. water absorption capacity	%	0.20	6.10	1.22	0.77
RCA water absorption capacity	%	1.93	18.91	5.38	2.33
Natural aggregate density	kg·m $^{-3}$	2482	2880	2616	84.67
RCA density	kg·m $^{-3}$	1800	2602	2312	121.88
Natural aggregate max. particle size	mm	8.00	32.00	20.00	3.80
RCA max. particle size	mm	8.00	32.00	18.95	4.76
Water content	kg·m $^{-3}$	108.30	234.00	170.69	31.55
28-day MOE (output)	GPa	11.30	54.80	30.41	7.81

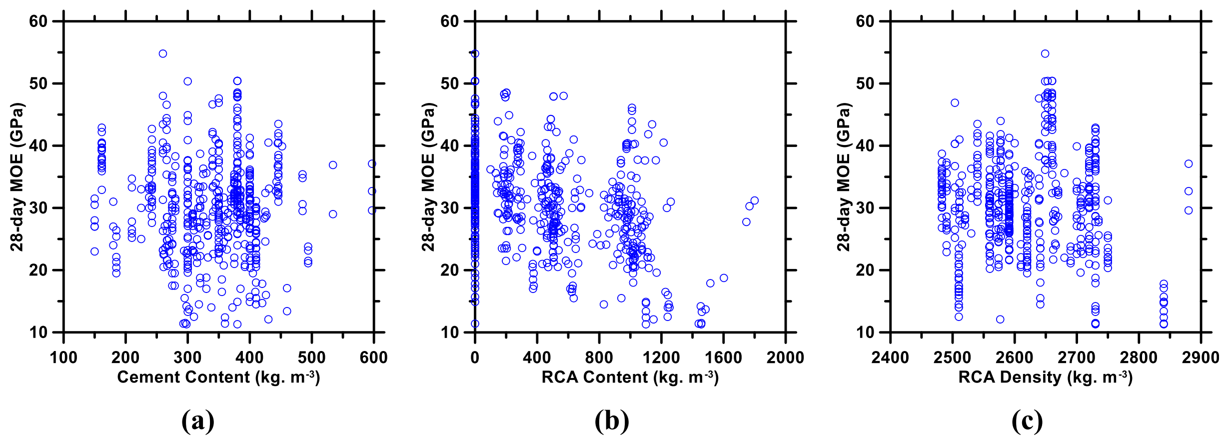

Fig. 1. Two-dimensional (2D) plots showing the 28-day MOE of RCA concrete in relation to the: (a) cement content; (b) RCA content; and (c) RCA density. In each plot, in addition to the one input parameter (in the x-axis) that is listed, there are variations in the 12 other input parameters, thus resulting in significant variation in the concrete' MOE.

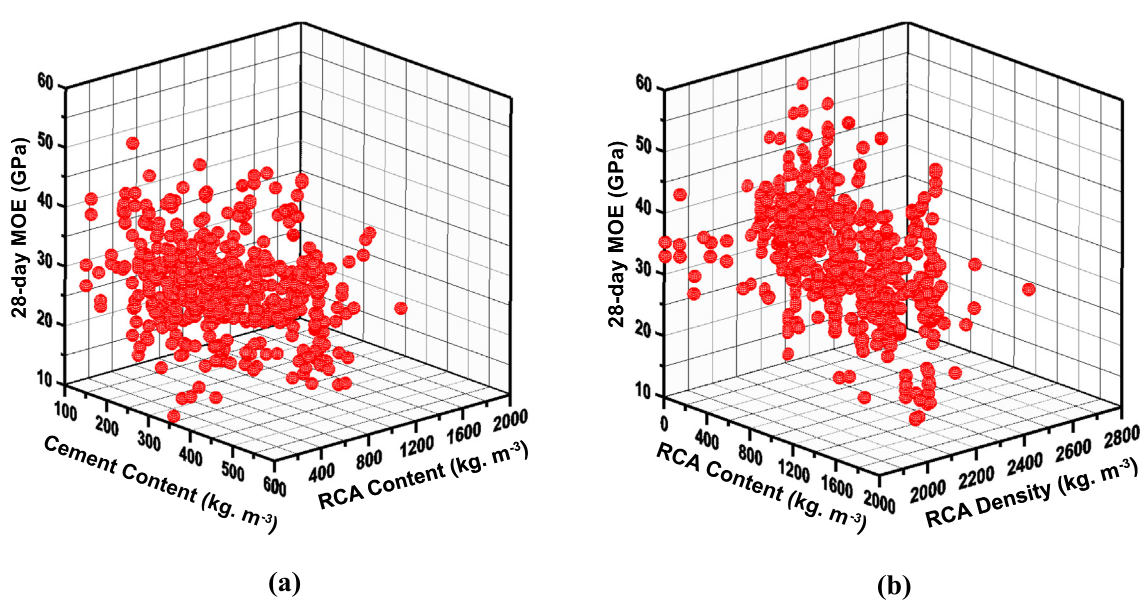

Fig. 2. 3D plots showing the 28-day MOE of RCA concrete in relation to the: (a) cement and RCA contents; and (b) RCA content and density. In each plot, in addition to the two input parameters (in the x- and y-axis) that are listed, there are variations in the other 11 other input parameters, thus resulting in significant variation in the concrete' MOE.

value) and the worst ML model would obtain a value of 1 (or the highest value). Therefore, on the sole the basis of CPI values – which accounts for all five statistical parameters (i.e., performance measures) – the ML models can be ranked (from best to worst) in terms of their prediction performances.

4. Results and discussion

4.1. Prediction of MOE of RCA concrete

As described previously in Section 3, each of the five standalone ML models and the ensemble ML model were firstly trained using 75% (selected randomly) of the database, and then the models' prediction performances were assessed against the remaining 25% of the database. Predictions of MOE of concrete from the testing set, as produced by the six ML models implemented in this study, are shown in Fig. 3; statistical parameters associated with models' prediction performances are enumerated in Table 2.

As can be seen in Fig. 3, all ML models produced predictions with reasonable accuracy, with the Pearson correlation coefficient (R) ranging from 0.67 to 0.93, and the root mean squared error (RMSE) ranging from 6.02 GPa to 2.93 GPa. Based on values of the latter, it can be said that the ML models can predict the MOE of RCA concrete formulated as per different mixture designs and contents and characteristics of RCA, within approximately ± 6 GPa and ± 3 GPa of the actual value in the worst- and best-case scenarios, respectively. Such margins of error in predictions are reasonable considering that even in experimental measurements of concretes' MOE, the standard deviation could be as high as 1–3 GPa [13,18]. On the basis of CPI values – the unified measure of accuracy of (or errors associated with) predictions – the prediction performance of the ML models can be ranked as ensemble ML model (voting: RF + SVM) > RF > MLP-ANN > GPR > LR > SVM. It is interesting to note that the SVM model has the poorest prediction performance, although several prior studies [26–28,68] have shown that the model can predict the compressive strength of conventional concretes with relatively high degree of accuracy (i.e., $R^2 \approx 0.90$). This implies that the composition-property correlations [i.e., links between inputs (concrete mixture design and RCA characteristics) and output (concretes' MOE)] in RCA concretes are far more complex than those in conventional concretes. It is conceivable that the additional complexity in the said composition-property correlations arise from the RCA – which, depending on its physiochemical characteristics, can drastically affect concrete's properties. The inferior prediction performance of the SVM model can also be explained on the bases of theories advanced in past studies [49,69,70]. These studies have reported that SVM models, very much like ANN models, rely on local search and optimization algorithms; as such, they suffer from the drawback of converging to a local minimum rather than the global minimum, especially when the relationship between input variables and output in the

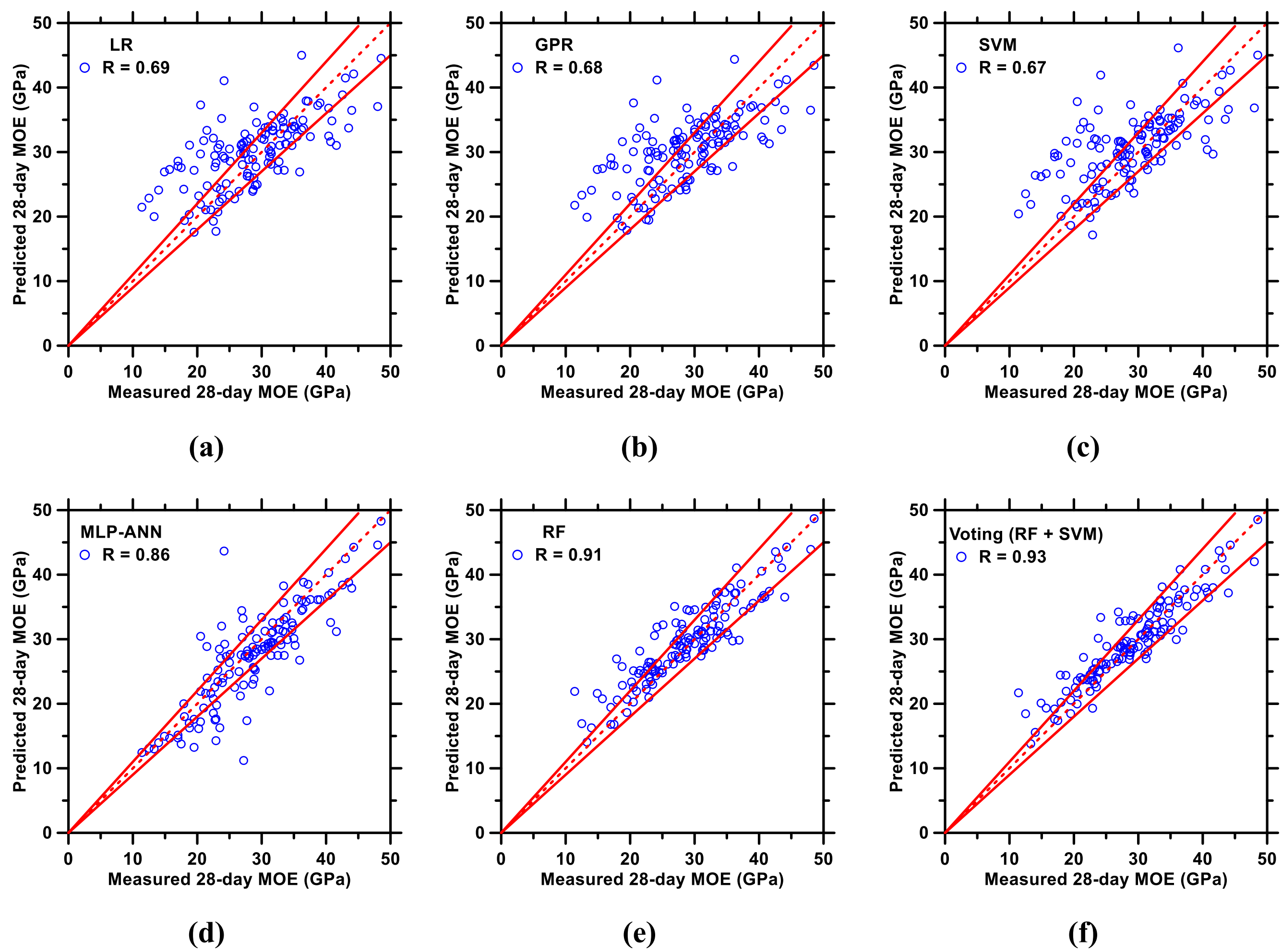

Fig. 3. Predictions of MOE of concrete produced by ML models: (a) linear regression (LR); (b) Gaussian process regression (GPR); (c) support vector machine (SVM); (d) multilayer perceptron artificial neural network (MLP-ANN); (e) random forests (RF); and (f) the ensemble (RF + SVM) ML model compared against measured values of the test set (comprising of 25% of the parent database). The dashed line represents the line of ideality, and the solid lines represent a ± 10% bound.

Table 2
Prediction performance of ML models, measured on the basis of the test set (i.e., 25% of the parent database). Five statistical parameters (i.e., R, R^2, MAE, MAPE, and RMSE) and the composite performance index (CPI) are shown. The best and the worst performing ML models are highlighted in bold.

ML Model	R *Unitless*	R^2 *Unitless*	MAE *GPa*	MAPE *%*	RMSE *GPa*	CPI *Unitless*
SVM	**0.6672**	**0.4452**	**4.4568**	**69.999**	**6.0226**	**0.9880**
LR	0.6901	0.4762	4.4064	69.208	5.7421	0.9296
GPR	0.6752	0.4559	4.5284	71.123	5.8773	0.9796
MLP	0.8559	0.7326	3.0973	48.646	4.3859	0.3770
RF	0.9119	0.8316	2.5198	39.577	3.3448	0.1241
Ensemble ML Model	**0.9332**	**0.8709**	**2.1469**	**33.719**	**2.9265**	**0.0000**

training dataset contains several closely-placed local minima. This, ultimately, manifests as poor prediction performance. Other studies have posited that this deficiency of the SVM model can be rectified by using Genetic programing [36,71] or bootstrap aggregation – for example, bagging, voting, grading, or stacking [26,72] – of outputs of one or more ML models in conjunction with output of the SVM model. This aspect – of improving prediction performance of the SVM model – has been examined in this study by combining it with the RF model within the ensemble (weighted voting scheme based) ML model. Further details pertaining to prediction performance of the ensemble ML model are provided later in this section.

Going back to Table 2, it can be seen that the LR and GPR models produced predictions of concretes' MOE with superior (albeit, only by a small margin) accuracy compared to the SVM model, but with comprehensively lower accuracy compared to the RF and MLP-ANN models. Such limitations in prediction performances of the LR and GPR models are not surprising because the former uses linear functions and the latter uses Gaussian distribution functions over datasets that – in all likelihood – feature far more complex correlations between the input variables and the output.

Predictions made by the MLP-ANN model were more accurate than all but one (i.e., the RF model) of the standalone ML models implemented in this study (Fig. 3; Table 2). As previously stated in Section 1, prediction performance of the MLP-ANN model could be compromised due to its inherent susceptibility to converge to a local – as opposed to the global – minimum [48,49]. However, in this study, the hyper-parameters (i.e., number of hidden layers and number of neurons per hidden layer) of the MLP-ANN model were rigorously optimized through the 10-fold CV method; on account of this optimization, it is expected that the aforementioned drawback of the model was overcome – at least partially – thereby allowing the model to produce predictions with reasonable accuracy. It is worth pointing out that in a recent study conducted by Sadati et al. [18], the MLP-ANN model was also used to predict concretes' MOE, whilst using a database that was similar to the one used in our study. As would be expected, due to these similarities, prediction performances of MLP-ANN models used in the two studies are similar (i.e., $R = 0.86$ in the current study and $R = 0.88$ in the study of Sadati et al. [18]). The minor difference, in prediction performances of the models as reported in the two studies, can be explained on account of differences in the following factors:

- Splitting of the parent database into training and testing sets: In the study of Sadati et al. [18], the parent database was split as per 90% and 10% between the training and test sets, respectively; in this study, a 75% and 25% split was used. The nature of split in the parent database affects both training and testing aspects of the model, which could ultimately affect its prediction performance [68].
- Number of input parameters: In the study of Sadati et al. [18], up to 21 input parameters – including some derivatives (i.e., derived from combination of two or more of the input parameters) – were used as inputs. In contrast, in this study, the database was pre-processed to ensure that none of the 13 input parameters were redundant.
- Procedure used for hyper-parameter optimization: In the study of Sadati et al., the number of hidden layers, as well as the number of neurons in each hidden layer, within the MLP-ANN model were varied manually, in an iterative manner, to determine the optimal values that produced the most accurate predictions (assessed in terms of the value of R). In this study, the 10-fold CV method, as described in Section 2, was used for optimization of both hyper-parameters of the MLP-ANN model.
- Training methodology: In the study reported in [18], the training dataset was split into multiple subsets; the splitting was done on the basis of similarities in RCA characteristics (i.e., maximum particulate size, density, and water absorption capacity). Each subset comprised of one reference system – that is, concrete devoid of any RCA – with its MOE normalized to 1.0, and the MOE of all RCA concrete mixtures in the subset were described as fractions of the MOE of the reference system. While this procedure enables classification of the training dataset into multiple subsets, which is potentially beneficial for training of the model, it could also introduce errors on account of creating *false equivalency* among reference systems of the different subsets [MOE of all reference systems, regardless of their mixture designs (e.g., contents of cement, water, and SCM), were labelled as 1.0 in the database]. In this study, to avoid the aforesaid *false equivalency*, absolute values of MOE (in the units of GPa) of each concrete specimen – regardless of its RCA content – were used in the training as well as the testing datasets.

The RF model outperformed all of the aforementioned standalone ML models in terms of prediction accuracy (Fig. 3; Table 2). This result is not surprising – and, in fact, in very good agreement with prior studies [27,49,73,74] that have also reported that the prediction performance of RF model often supersedes those of several standalone ML models. Superior prediction performance of the RF model is attributed to its structure – which comprises of a large number of "deep" trees that are grown without any smoothening or pruning. The unpruned "deep" trees allow data in the training set to be split in a logical manner, which, in turn, not only reduces generalization errors but also serves to mitigate overfitting (high bias) of the training data. Furthermore, the two-stage randomization employed in the RF model (see **Supplementary Information** for more details) diminishes correlation among the unpruned trees, thus reducing variation (underfitting) and ensuring homogeneity among the data represented in each of the tree-nodes. It is interesting to note that prediction performance of the RF model was bolstered when it was hybridized with the SVM model to form the ensemble, voting-based ML model. As shown in Fig. 3, for majority of concretes in the test set, the ensemble ML model was able to predict MOE within ±10% of the actual values; even when the predictions fell outside of the ±10% bound, the RMSE of MOE predictions was still very low (i.e., <3 GPa). This improvement in the accuracy of predictions of the ensemble ML model, vis-à-vis the standalone SVM and RF, is credited to the ensemble model's ability to combine predictions of the SVM and RF model, in a meta-heuristic manner, such that each prediction is rendered more accurate than the one yielded by either of two standalone ML models. In order words, the ensemble ML model produces accurate predictions by ensuring that prediction errors made by one model, in a given subset of the data, are compensated by superior prediction accuracy of the other model.

4.2. Optimization of mixture design of RCA concrete

The results shown in Section 4, and the accompanying discussions, show that the ensemble ML model – comprised of SVM and RF models within a weighted voting scheme – is able to reliably predict the MOE of RCA concretes in relation to their mixture design and RCA characteristics. In this section, the training of the ensemble ML model (on the basis of which the model learned the functional input(s)-output correlations), and its ability to dependably predict MOE of concrete, were leveraged to develop an optimization module. The objective of the optimization module is to predict optimal mixture design of RCA concrete to achieve target (i.e., user-imposed) value(s) of 28-day MOE. The optimization module was designed to accept the target 28-day MOE as an input along with selected inputs of concrete mixture design and RCA characteristics. The module, then, performed predictions in reverse – using a Bayesian optimization approach [75,76], while leveraging the input–output correlation(s) learned by the ensemble ML model during the training process – to reveal optimal values of the remaining concrete mixture design parameters that would produce the target 28-day MOE.

Two different scenarios were used for MOE optimization – one for RCA concrete formulated using 380 $kg{\cdot}m^{-3}$ of plain (OPC) binder, and the other for concrete formulated using 380 $kg{\cdot}m^{-3}$ of binary/ternary binders (prepared by replacing up to 50% of OPC with fly ash and/or blast furnace slag). In both scenarios, certain input parameters were fixed: total coarse aggregate content (1000 $kg{\cdot}m^{-3}$); fine aggregate content (642 $kg{\cdot}m^{-3}$); natural coarse aggregate water absorption capacity (0.8%); natural coarse aggregate density (2590 $kg{\cdot}m^{-3}$); RCA density (2270 $kg{\cdot}m^{-3}$); natural coarse aggregate nominal maximum particle size (20 mm); and RCA nominal maximum particle size (20 mm). These values were chosen as they had the highest frequency (of appearance) in the database. Next, in both scenarios, the RCA content in the concrete was varied by replacing 0–100% of the natural coarse aggregate, while keeping the total coarse aggregate content fixed at 1000 $kg{\cdot}m^{-3}$. Furthermore, in both scenarios, two different values of RCA water absorption capacity were used: a low value of 5.3% and a high value of 10%. Here, it is acknowledged that the lower bound of RCA water absorption capacity (i.e., 5.3%) is rather larger than the values that have been reported in past literature [7,12,17,19]. However, this value was chosen because of its higher frequency in the training database (compared to values $\ll$5.3%); this is important to ensure that the ML model is adequately trained and, thus, able to produce accurate results during the optimization routines. In the last step, the optimization module was used to determine the optimum contents of water (and, hence, the optimum values of w/cm) that ought to be used for concrete formulation so as to achieve target MOE of 30, 40, or 50 GPa in relation to the content and water absorption capacity of RCA in the concrete. Further details pertaining to the (fixed and variable) input parameters that were used in the optimizations are listed in Table 3; results produced by the optimization module are illustrated in Fig. 4.

Table 3
Parameters used for optimization of mixture design of RCA concretes. Of the 13 parameters, 12 were used as inputs, and 1 (i.e., water content) was calculated by the ensemble ML model as output. The parameters highlighted in bold were used as inputs, but varied between different optimizations.

Attribute	Unit	Scenario 1	Scenario 2
Cement (OPC) content	$kg{\cdot}m^{-3}$	380	300
SCM (fly ash and/or slag) content	$kg{\cdot}m^{-3}$	0	80
Total coarse aggregate content	$kg{\cdot}m^{-3}$	1000	1000
RCA content	$kg{\cdot}m^{-3}$	0–100% of 1000	0–100% of 1000
Fine aggregate content	$kg{\cdot}m^{-3}$	642	642
RCA water absorption capacity	%	5.3% (low) or 10% (high)	5.3% (low) or 10% (high)
Natural agg. water absorption capacity	%	0.8	0.8
Natural aggregate density	$kg{\cdot}m^{-3}$	2590	2590
RCA density	$kg{\cdot}m^{-3}$	2270	2270
Natural aggregate max. particle size	mm	20	20
RCA max. particle size	mm	20	20
28-day MOE	GPa	30–50 GPa	30–50 GPa
Water content	$kg{\cdot}m^{-3}$	Output	Output

As shown in Fig. 4, for all concrete mixtures, the optimum w/cm (calculated based on the optimum water content, produced as output of the optimization module) decreases with increasing RCA content in the concretes. Notably, this trend – decrease in w/cm with respect to increasing RCA content – is more pronounced in plain OPC concrete than in binary/ternary binders-based concrete. The aforesaid trend is expected because at higher RCA content, owing to the intrinsic deficiencies of RCA (e.g., weak aggregate-mortar interface, microcracks, and enlargement of ITZ volume [18,77,78]), the MOE of the host material (i.e., RCA concrete) progressively reduces. To compensate for this RCA-induced reduction of MOE, the water content in the concrete – and, thus, its porosity – ought to be reduced. Fig. 4 also shows that for any given concrete mixture design – regardless of the content and water absorption capacity of RCA, or the presence/absence of SCMs – to enhance MOE (e.g., for 30 to 50 GPa), the w/cm should be reduced. This trend, again, is stimulated by the porosity; low w/cm ensures low porosity, thereby enhancing the solid to solid phase connectivity within the concrete microstructure, which, in turn, enhances the concrete's MOE. Lastly, results from the optimizations show that for any concrete mixture design, regardless of the RCA content, a lower w/cm is needed when RCA with higher water absorption capacity is used to secure a given MOE. This, akin to the trends described above, suggests that reducing concrete's porosity (by using lower w/cm) can effectively compensate for the inherent deficiencies of the RCA – in this case, its higher water absorption capacity [79].

It should be clarified that results shown in Fig. 4 strictly pertain to concrete formulated using mixture design parameters and RCA characteristics listed in Table 3. As such, the results merely serve as examples to exhibit the versatility of the ensemble ML model – in particular, its ability to go beyond predictions and perform optimizations of concrete mixture designs based on desired performance criterion. Nevertheless, and as elucidated in the above discussion, the general trends that emerged in Fig. 4 are fully compliant with theoretical laws, and are in good agreement with results reported in prior studies [18,78–81]. These results, therefore, validate predictions and optimizations produced by the ensemble ML model. More importantly, the result highlight that with proper and rigorous training of ML models, rapid and reliable predictions of RCA concretes' properties and optimization of their mixture designs are indeed feasible.

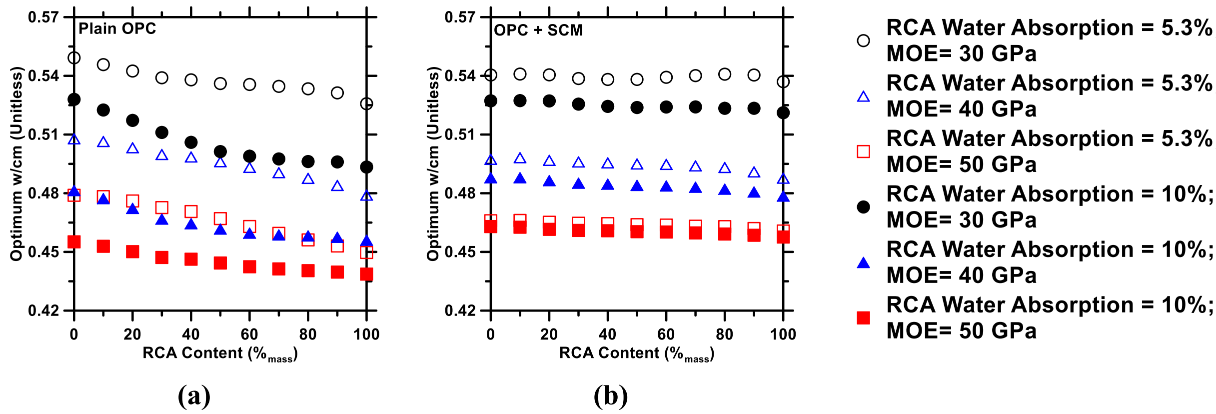

Fig. 4. Optimal values of water to cementitious materials mass ratio (w/cm) – produced by the ensemble ML model to achieve target MOE of 30 GPa, or 40 GPa, or 50 GPa – plotted against the RCA content in concrete mixtures formulated using: (a) plain OPC binder (scenario 1); and (b) binary/ternary binder, wherein OPC is partially substituted with SCMs (fly ash, or blast furnace slag, or a combination of both). RCA was used to replace 0–$100\%_{mass}$ of natural coarse aggregate in the concrete. Hollow and solid symbols represent cases wherein low (5.3%) and high (10%) values of RCA water absorption capacity, respectively, were used. Other fixed and variable input parameters, used in the optimizations, are listed in Table 3.

5. Conclusion

This study presented the first application of an ensemble machine learning (ML) model – formulated by combining the random forests (RF) model with support vector machine (SVM) within a weighted voting framework – to predict the modulus of elasticity (MOE) of recycled aggregate concrete [i.e., wherein the natural coarse aggregate component of concrete is partially or fully replaced by coarse recycled concrete aggregate (RCA)].

The ensemble ML model was trained using 75% of a parent database, comprising of >500 experimentally-obtained, distinct data-records. Each data-record featured 13 inputs (i.e., variables pertaining to concrete mixture design and physiochemical characteristics of aggregates) and 1 output (28-day MOE of concrete). After the model's training, input–output correlations learned during the training process were utilized by the model to enable predictions in blank, untrained data-domains (i.e., remaining 25% of the parent database). The hybrid model's prediction performance was measured using 5 different statistical parameters and a composite performance index (CPI), and compared against 5 standalone ML models (i.e., LR, GPR, MLP-ANN, SVM, and RF).

The MLP-ANN model was able to predict flotation outcomes with reasonable accuracy – with prediction performances that were better than SVM model and other ML models premised on linear input–output functions (i.e., LR and ENR models). However, owing to the inability of MLP-ANN to converge to the global minimum, its prediction performance was secondary compared to the standalone RF and ensemble ML models. Excellent prediction performance of the RF model was attributed to its structure – which comprises of a large number of "deep" trees that are grown without any smoothening or pruning. The prediction performance of the RF model was further bolstered when it was combined with SVM. Such improvement in prediction performance of the ensemble ML model was attributed to its ability to combine predictions of the SVM and RF model, in a metaheuristic manner, thus rendering each prediction more accurate than the one yielded by either of two standalone ML models.

Next, the training of the ensemble ML model, and its ability to reliably predict MOE of concrete, were leveraged to develop an optimization module to predict optimal mixture designs of RCA concretes that are expected to achieve target (i.e., user-imposed) value(s) of 28-day MOE. Results obtained from such optimizations showed that in RCA concrete, high MOE (up to 50 GPa at 28 days) can be achieved provided that the water to cementitious materials ratio (w/cm) is reduced (within the range of 0.43–0.55) in relation to: (i) increasing RCA content; (ii) increasing water absorption capacity of the RCA; and (iii) increasing value of the target MOE.

The excellent prediction performance (i.e., RMSE of ≈ 3.0 GPa) of the ensemble ML model, and its ability to perform optimizations that are in alignment with theoretical laws, indicates that the model could serve as a reliable tool to promptly and accurately predict and optimize concrete's properties in relation to its mixture design. While only 526 data-records were used in this study for training and testing of the ML model, it is expected that enlargement of the database volume would further improve the accuracy and reliability of predictions and optimizations. The authors envision that the utilization of a superior, more diverse database – wherein additional influential physiochemical characteristics of concrete (e.g., particle size distribution; Los Angeles abrasion loss of RCA; and cement's, SCM's and aggregate's chemical composition) are properly represented – will elicit even further improvements in prediction and optimization performance of the ensemble ML model.

Declaration of Competing Interest

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

Acknowledgements

Experimental and computational tasks, described in this study, were conducted in the Advanced Construction Materials Laboratory (AMCL) and the Materials Research Center (MRC), respectively, of Missouri S&T. Funding for this study was provided by the Leonard Wood Institute (LWI), RE-CAST Tier-1 University Transportation Center at Missouri S&T, and the National Science Foundation (NSF; CMMI: 1661609 and CMMI: 1932690). The authors would like to thank Dr. Seyedhamed Sadati for his help with mining and organization of the database.

Appendix:. Research data (Mendeley)

Detailed descriptions of standalone and ensemble ML models presented in this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.conbuildmat.2020.118271.

References

[1] O. US EPA, Construction and Demolition: Material-Specific Data, US EPA. (2017). https://www.epa.gov/facts-and-figures-about-materials-waste-and-recycling/construction-and-demolition-material-specific (accessed 02.05.19).
[2] O. US EPA, Resource Conservation and Recovery Act (RCRA) in Focus: Hazardous Waste Generator Guidance by Industry, US EPA. (2015). https://www.epa.gov/hwgenerators/resource-conservation-and-recovery-act-rcra-focus-hazardous-waste-generator-guidance (accessed 26.04.19).
[3] O. US EPA, Sustainable Management of Construction and Demolition Materials, US EPA. (2016). https://www.epa.gov/smm/sustainable-management-construction-and-demolition-materials (accessed 02.05.19).
[4] O. Isgor, Cracking susceptibility of concrete made with recycled concrete, Aggregate (2013).
[5] A. Behnood, J. Olek, M.A. Glinicki, Predicting modulus elasticity of recycled aggregate concrete using M5′ model tree algorithm, Constr. Build. Mater. 94 (2015) 137–147, https://doi.org/10.1016/j.conbuildmat.2015.06.055.
[6] A. Abbas, G. Fathifazl, O.B. Isgor, A.G. Razaqpur, B. Fournier, S. Foo, Environmental Benefits of Green Concrete, in: 2006 IEEE EIC Climate Change Conference, IEEE, Ottawa, ON, Canada, 2006: pp. 1–8. Doi: 10.1109/EICCCC.2006.277204.
[7] J.-Zh. Xiao, J.-B. Li, Ch. Zhang, On relationships between the mechanical properties of recycled aggregate concrete: aAn overview, Mater. Struct. 39 (2007) 655–664, https://doi.org/10.1617/s11527-006-9093-0.
[8] R. Feiz, J. Ammenberg, L. Baas, M. Eklund, A. Helgstrand, R. Marshall, Improving the CO 2 performance of cement, part I: utilizing life-cycle assessment and key performance indicators to assess development within the cement industry, J. Cleaner Prod. 98 (2015) 272–281.
[9] M. Schneider, Process technology for efficient and sustainable cement production, Cem. Concr. Res. 78 (2015) 14–23, https://doi.org/10.1016/j.cemconres.2015.05.014. Part A.
[10] J.J. Biernacki, J.W. Bullard, G. Sant, K. Brown, F.P. Glasser, S. Jones, T. Ley, R. Livingston, L. Nicoleau, J. Olek, others, Cements in the 21st century: challenges, perspectives, and opportunities, J. Am. Ceram. Soc. 100 (2017) 2746–2773, https://doi.org/10.1111/jace.14948.
[11] C. Li, H. Geng, C. Deng, B. Li, S. Zhao, Experimental investigation on columns of steel fiber reinforced concrete with recycled aggregates under large eccentric compression load, Materials 12 (2019) 445, https://doi.org/10.3390/ma12030445.
[12] V. Corinaldesi, Mechanical and elastic behaviour of concretes made of recycled-concrete coarse aggregates, Constr. Build. Mater. 24 (2010) 1616–1620, https://doi.org/10.1016/j.conbuildmat.2010.02.031.
[13] T.C. Hansen, E. Boegh, Elasticity and drying shrinkage concrete of recycled-aggregate, J. Proceed. (1985) 648–652.
[14] L. Evangelista, J. De Brito, Durability performance of concrete made with fine recycled concrete aggregates, Cem. Concr. Compos. 32 (2010) 9–14.
[15] N.F. Günçan, Using waste concrete as aggregate, Cem. Concr. Res. 25 (1995) 1385–1390.
[16] M. Etxeberria, E. Vázquez, A. Marí, M. Barra, Influence of amount of recycled coarse aggregates and production process on properties of recycled aggregate concrete, Cem. Concr. Res. 37 (2007) 735–742.
[17] J. Xiao, J. Li, C. Zhang, Mechanical properties of recycled aggregate concrete under uniaxial loading, Cem. Concr. Res. 35 (2005) 1187–1194.
[18] S. Sadati, L.E. Brito da Silva, D.C. Wunsch, K.H. Khayat, Artificial intelligence to investigate modulus of elasticity of recycled aggregate concrete, ACI Mater. J. 116 (2019), https://doi.org/10.14359/51706948.
[19] W. Liu, W. Cao, J. Zhang, R. Wang, L. Ren, Mechanical behavior of recycled aggregate concrete-filled steel tubular columns before and after fire, Materials 10 (2017) 274, https://doi.org/10.3390/ma10030274.
[20] P. Folino, H. Xargay, Recycled aggregate concrete – mechanical behavior under uniaxial and triaxial compression, Constr. Build. Mater. 56 (2014) 21–31, https://doi.org/10.1016/j.conbuildmat.2014.01.073.
[21] M. Casuccio, M.C. Torrijos, G. Giaccio, R. Zerbino, Failure mechanism of recycled aggregate concrete, Constr. Build. Mater. 22 (2008) 1500–1506, https://doi.org/10.1016/j.conbuildmat.2007.03.032.
[22] A.C.I. Committee 318, A.C. Institute, I.O. for Standardization, Building Code Requirements for Structural Concrete (ACI 318-08) and Commentary, American Concrete Institute, 2008. https://books.google.com/books?id=c6yQszMV2-EC.
[23] R. Sri Ravindrarajah, C.T. Tam, Properties of concrete made with crushed concrete as coarse aggregate, Mag. Concr. Res. 37 (1985) 29–38.
[24] C.T. Tam, T.Y. Lim, R. Sri Ravindrarajah, S.L. Lee, Relationship between strength and volumetric composition of moist-cured cellular concrete, Mag. Concr. Res. 39 (1987) 12–18.
[25] R.K. Dhir, M.C. Limbachiya, T. Leelawat, Suitability of recycled concrete aggregate for use in BS 5328 designated mixes, in: Proceedings of the Institution of Civil Engineers: Structures and Buildings, 1999.
[26] J.-S. Chou, C.-F. Tsai, A.-D. Pham, Y.-H. Lu, Machine learning in concrete strength simulations: multi-nation data analytics, Constr. Build. Mater. 73 (2014) 771–780, https://doi.org/10.1016/j.conbuildmat.2014.09.054.
[27] B.A. Young, A. Hall, L. Pilon, P. Gupta, G. Sant, Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods, Cem. Concr. Res. 115 (2019) 379–388.
[28] K.O. Akande, T.O. Owolabi, S. Twaha, S.O. Olatunji, Performance comparison of SVM and ANN in predicting compressive strength of concrete, IOSR J. Comput. Eng. 16 (2014) 88–94.
[29] A. Behnood, V. Behnood, M.M. Gharehveran, K.E. Alyamac, Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm, Constr. Build. Mater. 142 (2017) 199–207.
[30] J.-S. Chou, C.-K. Chiu, M. Farfoura, I. Al-Taharwa, Optimizing the prediction accuracy of concrete compressive strength based on a comparison of data-mining techniques, J. Comput. Civil Eng. 25 (2010) 242–253.
[31] Z.-H. Duan, S.-C. Kou, C.-S. Poon, Prediction of compressive strength of recycled aggregate concrete using artificial neural networks, Constr. Build. Mater. 40 (2013) 1200–1206.
[32] R. Gupta, M.A. Kewalramani, A. Goel, Prediction of concrete strength using neural-expert system, J. Mater. Civ. Eng. 18 (2006) 462–466.
[33] J. Kasperkiewicz, J. Racz, A. Dubrawski, HPC strength prediction using artificial neural network, J. Comput. Civil Eng. 9 (1995) 279–284.
[34] N.K. Nagwani, S.V. Deo, Estimating the concrete compressive strength using hard clustering and fuzzy clustering based regression techniques, Sci. World J. 2014 (2014).
[35] B.A. Omran, Q. Chen, R. Jin, Comparison of data mining techniques for predicting compressive strength of environmentally friendly concrete, J. Comput. Civil Eng. 30 (2016) 04016029.
[36] V. Veloso de Melo, W. Banzhaf, Improving the prediction of material properties of concrete using Kaizen programming with simulated annealing, Neurocomputing. 246 (2017) 25–44, https://doi.org/10.1016/j.neucom.2016.12.077.
[37] I.-C. Yeh, Modeling of strength of high-performance concrete using artificial neural networks, Cem. Concr. Res. 28 (1998) 1797–1808.
[38] I.-C. Yeh, Modeling concrete strength with augment-neuron networks, J. Mater. Civ. Eng. 10 (1998) 263–268.
[39] I.-C. Yeh, L.-C. Lien, Knowledge discovery of concrete material using genetic operation trees, Expert Syst. Appl. 36 (2009) 5807–5812.
[40] M.F. Zarandi, I.B. Türksen, J. Sobhani, A.A. Ramezanianpour, Fuzzy polynomial neural networks for approximation of the compressive strength of concrete, Appl. Soft Comput. 8 (2008) 488–498.
[41] Z.H. Duan, S.C. Kou, C.S. Poon, Using artificial neural networks for predicting the elastic modulus of recycled aggregate concrete, Constr. Build. Mater. 44 (2013) 524–532, https://doi.org/10.1016/j.conbuildmat.2013.02.064.
[42] A. Gholampour, I. Mansouri, O. Kisi, T. Ozbakkaloglu, 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. Appl. (2018), https://doi.org/10.1007/s00521-018-3630-y.
[43] N. Deshpande, S. Londhe, S. Kulkarni, Modeling compressive strength of recycled aggregate concrete by artificial neural network, model tree and non-linear regression, Int. J. Sustain. Built Environ. 3 (2014) 187–198, https://doi.org/10.1016/j.ijsbe.2014.12.002.
[44] E. Frank, M. Hall, L. Trigg, G. Holmes, I.H. Witten, Data mining in bioinformatics using Weka, Bioinformatics 20 (2004) 2479–2481.
[45] F. Deng, Y. He, S. Zhou, Y. Yu, H. Cheng, X. Wu, Compressive strength prediction of recycled concrete based on deep learning, 175 (2018) 562–569. https://doi.org/10.1016/j.conbuildmat.2018.04.169.
[46] P. Cunningham, J. Carney, S. Jacob, Stability problems with artificial neural networks and the ensemble solution, Artif. Intell. Med. 20 (2000) 217–225.
[47] X. Yao, Evolving artificial neural networks, Proc. IEEE 87 (1999) 1423–1447.
[48] G. Zhang, B.E. Patuwo, M.Y. Hu, Forecasting with artificial neural networks: the state of the art, Int. J. Forecast. 14 (1998) 35–62.
[49] R. Cook, J. Lapeyre, H. Ma, A. Kumar, Prediction of compressive strength of concrete: a critical comparison of performance of a hybrid machine learning model with standalone models, ASCE J. Mater. Civil Eng. 31 (2019) 04019255, https://doi.org/10.1061/(ASCE)MT.1943-5533.0002902.
[50] J.R. Quinlan, Learning with Continuous Classes, in: Proceedings of the Australian Joint Conference on Artificial Intelligence, World Scientific, Singapore, 1992: pp. 343–348. https://researchcommons.waikato.ac.nz/handle/10289/1183 (accessed 05.12.18).
[51] Y. Wang, I.H. Witten, Induction of model trees for predicting continuous classes, in: Proceedings of European Conference on Machine Learning, University of Economics, Faculty of Informatics and Statistics, Prague, 1997. https://researchcommons.waikato.ac.nz/handle/10289/1183 (accessed 05.12.18).
[52] L. Breiman, Random forests, Machine Learning. 45 (2001) 5–32.
[53] C.-C. Chia, I. Rubinfeld, B.M. Scirica, S. McMillan, H.S. Gurm, Z. Syed, Looking Beyond Historical Patient Outcomes to Improve Clinical Models, Science Translational Medicine. 4 (2012) 131ra49–131ra49. Doi: 10.1126/scitranslmed.3003561.
[54] R.J. Schalkoff, Artificial neural networks, McGraw-Hill, New York, 1997.
[55] M.W. Gardner, S.R. Dorling, Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences, Atmos. Environ. 32 (1998) 2627–2636, https://doi.org/10.1016/S1352-2310(97)00447-0.

[56] I.H. Witten, E. Frank, M.A. Hall, Data mining: practical machine learning tools and techniques, 3rd ed., Morgan Kaufmann, Burlington, MA, 2011.
[57] O. Bousquet, U. von Luxburg, G. Rätsch, eds., Advanced lectures on machine learning: ML Summer Schools 2003, Canberra, Australia, February 2-14, 2003 [and] Tübingen, Germany, August 4-16, 2003: revised lectures, Springer, Berlin ; New York, 2004.
[58] J. Bernardo, J. Berger, A. Dawid, A. Smith, others, Regression and classification using Gaussian process priors, Bayesian Stat. 6 (1998) 475.
[59] C. Strobl, A.-L. Boulesteix, A. Zeileis, T. Hothorn, Bias in random forest variable importance measures: Illustrations, sources and a solution, BMC Bioinf. 8 (2007) 25, https://doi.org/10.1186/1471-2105-8-25.
[60] K.J. Archer, R.V. Kimes, Empirical characterization of random forest variable importance measures, Comput. Stat. Data Anal. 52 (2008) 2249–2260, https://doi.org/10.1016/j.csda.2007.08.015.
[61] R. Polikar, Ensemble learning, in: Ensemble Machine Learning, Springer, 2012: pp. 1–34.
[62] T.G. Dietterich, Ensemble methods in machine learning, in: International Workshop on Multiple Classifier Systems, Springer, 2000: pp. 1–15.
[63] C. Schaffer, Selecting a classification method by cross-validation, Machine Learning. 13 (1993) 135–143.
[64] C09 Committee, Test Method for Slump of Hydraulic-Cement Concrete, ASTM International, n.d. Doi: 10.1520/C0143_C0143M-15A.
[65] C09 Committee, Test Method for Air Content of Freshly Mixed Concrete by the Pressure Method, ASTM International, n.d. Doi: 10.1520/C0231_C0231M-17A.
[66] C09 Committee, Test Method for Static Modulus of Elasticity and Poissons Ratio of Concrete in Compression, ASTM International, n.d. Doi: 10.1520/C0469_C0469M-14.
[67] V. Chandwani, V. Agrawal, R. Nagar, Modeling slump of ready mix concrete using genetic algorithms assisted training of Artificial Neural Networks, Expert Syst. Appl. 42 (2015) 885–893.
[68] S.B. Kotsiantis, I.D. Zaharakis, P.E. Pintelas, Machine learning: a review of classification and combining techniques, Artif Intell Rev. 26 (2006) 159–190, https://doi.org/10.1007/s10462-007-9052-3.
[69] S.M. Clarke, J.H. Griebsch, T.W. Simpson, Analysis of support vector regression for approximation of complex engineering analyses, J. Mech. Des. 127 (2004) 1077–1087, https://doi.org/10.1115/1.1897403.
[70] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, B. Scholkopf, Support vector machines, IEEE Intell. Syst. Their Appl. 13 (1998) 18–28.
[71] P. Chopra, R.K. Sharma, M. Kumar, Prediction of compressive strength of concrete using artificial neural network and genetic programming, Adv. Mater. Sci. Eng. 2016 (2016), https://doi.org/10.1155/2016/7648467.
[72] R. Polikar, Ensemble based systems in decision making, IEEE Circuits Syst. Mag. 6 (2006) 21–45.
[73] N.M. Anoop Krishnan, S. Mangalathu, M.M. Smedskjaer, A. Tandia, H. Burton, M. Bauchy, Predicting the dissolution kinetics of silicate glasses using machine learning, J. Non-Cryst. Solids. 487 (2018) 37–45, https://doi.org/10.1016/j.jnoncrysol.2018.02.023.
[74] H. Liu, Z. Fu, K. Yang, X. Xu, M, Bauchy, Machine learning for glass science and engineering: A review, J. Non-Cryst. Solids 119419 (2019), https://doi.org/10.1016/j.jnoncrysol.2019.04.039.
[75] M. Pelikan, Hierarchical Bayesian Optimization Algorithm, in: M. Pelikan (Ed.), Hierarchical Bayesian Optimization Algorithm: Toward a New Generation of Evolutionary Algorithms, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005: pp. 105–129. Doi: 10.1007/978-3-540-32373-0_6.
[76] K. Swersky, J. Snoek, R.P. Adams, Multi-Task Bayesian Optimization, in: C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 26, Curran Associates, Inc., 2013: pp. 2004–2012. http://papers.nips.cc/paper/5086-multi-task-bayesian-optimization.pdf (accessed 19.06.18).
[77] A.M. Knaack, Y.C. Kurama, Behavior of reinforced concrete beams with recycled concrete coarse aggregates, J. Struct. Eng. 141 (2015) B4014009, https://doi.org/10.1061/(ASCE)ST.1943-541X.0001118.
[78] G. Andreu, E. Miren, Experimental analysis of properties of high performance recycled aggregate concrete, Constr. Build. Mater. 52 (2014) 227–235, https://doi.org/10.1016/j.conbuildmat.2013.11.054.
[79] S. Sadati, K.H. Khayat, Can concrete containing high-volume recycled concrete aggregate be durable?, ACI Mater J. Farmington Hills. 115 (2018) 471–480, https://doi.org/10.14359/51702190.
[80] N. Fonseca, J. de Brito, L. Evangelista, The influence of curing conditions on the mechanical performance of concrete made with recycled concrete waste, Cem. Concr. Compos. 33 (2011) 637–643, https://doi.org/10.1016/j.cemconcomp.2011.04.002.
[81] H. Cui, X. Shi, S. Memon, F. Xing, W. Tang, Experimental study on the influence of water absorption of recycled coarse aggregates on properties of the resulting concretes, J. Mater. Civ. Eng. 27 (2014) 04014138, https://doi.org/10.1061/(ASCE)MT.1943-5533.0001086.