Construction and Building Materials 492 (2025) 143126

Contents lists available at ScienceDirect

Construction and Building Materials

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

Machine learning approaches for RCPT modeling of concrete

Hamed Naseri [a,*], Farzad Safi Jahanshahi [b,1], Amir H. Gandomi [c]

[a] Department of Civil, Geological, and Mining Engineering, Polytechnique Montréal, Montréal, Canada
[b] Department of Civil Engineering, Sirjan University of Technology, Sirjan, Iran
[c] Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, Australia

A R T I C L E I N F O

Keywords:
Concrete
Rapid chloride permeability test
Ensemble learning
Self-tuning multilayer
SHapley Additive explanation

A B S T R A C T

Chloride ion penetration can lead to the corrosion of embedded steel reinforcement in concrete and deteriorate its durability. One of the commonly used methods for determining the durability of concrete is the rapid chloride permeability test (RCPT). The accurate prediction of RCPT is essential in optimizing concrete mixture design. In this study, RCPT modeling was conducted using a dataset of 469 samples. The contributions of this study are to apply a large-scale dataset with essential variables, to identify the optimal variables to maximize the performance of RCPT prediction models, to determine the most accurate prediction method for RCPT prediction, and to introduce a novel algorithm to tune hyperparameters of machine learning methods. To this end, first, feature selection was implemented to select the optimal variables. Then, different machine learning methods were used to predict RCPT. A novel less-parameter algorithm (STML) was adjusted to tune hyperparameters considering multiple metrics, which was considerably more accurate than the conventional tuning method. The results suggested that eXtreme Gradient Boosting tuned by STML was the best-performing model, with an MAE of 38.893 coulombs. Subsequently, SHapley Additive exPlanation was synchronized with the best-performing model, and the results showed that the test temperature had the highest relative influence on RCPT, followed by fly ash to binder ratio, silica fume to cement ratio, and coarse aggregate content. Finally, the Partial Dependence Plots were applied to capture the influence direction of different variables on RCPT, allowing the identification of the optimal range of materials to minimize RCPT.

1. Introduction

Concrete has been long-established in the construction industry due to its durability, fire resistance, affordability, ease of use, and versatility in taking on diverse shapes [1]. Currently, more than 10 billion tons of concrete are produced annually [2], and this volume will have reached about 18 billion tons by 2050 [3]. The manufacturing of concrete contributes significantly to environmental problems and greenhouse gas emissions. Each ton of cement production, as a significant component of concrete production, requires about 1.5 tons of raw materials, 4000 MJ of energy, and 140 kWh of electricity [4]. Further, 5–7 % of global CO_2 emissions are the contribution of ordinary Portland cement (OPC) production [5]. These environmentally harmful effects can grow due to inadequate durability since corrective actions, such as repair or replacement produce CO_2 emissions, use energy, and raw materials [6]. Therefore, to reduce such harmful effects and boost sustainability, the durability of concrete is obliged to be improved.

Reinforcing steel corrosion is a major factor restraining the durability of concrete structures. De-passivation of the embedded steel, a consequence of chloride ion contamination and carbonation, leads to this corrosion, which is known as a foremost challenge of reinforced structures [7]. Durability concerns in structural concrete arise from the transportation of chloride ions and subsequent reinforcement corrosion. While there is a well-established theory for determining ion diffusivity in solution based on Fick's laws, measuring chloride ion transport through concrete pores is challenging [8]. In other words, time and cost constraints make it difficult to determine chloride diffusion coefficients [9]. Additionally, achieving adequate chloride profiles within the specified testing time has proven challenging in producing high-quality concrete.

Various approaches have been applied to investigate the chloride ingress in concrete and repair materials. For example, Andrade et al. [10] developed an electrical resistivity-based model incorporating a reaction factor to account for chloride binding effects, significantly improving the prediction of effective chloride diffusion coefficients compared to traditional resistivity methods. Shi et al. [11] evaluated

* Corresponding author.
E-mail addresses: hamed.naseri@polymtl.ca (H. Naseri), farzad_jahanshahi@yahoo.com (F. Safi Jahanshahi), amirhossein.gandomi@uts.edu.au (A.H. Gandomi).
[1] https://orcid.org/0000-0002-9205-3497

https://doi.org/10.1016/j.conbuildmat.2025.143126
Received 21 January 2025; Received in revised form 19 July 2025; Accepted 8 August 2025
Available online 11 August 2025

Nomenclature

RCPT	Rapid chloride permeability test
MLP	Multi-Layer Perceptron
C	Cement content
S	Sand
FA	Fly ash content
SF	Silica fume content
CAg	Coarse aggregate content
FAg	Fine aggregate content
W	Water content
B	Binder content
T	Temperature
MK	Metakaolin
SP	Superplasticizer
OPC	Ordinary Portland Cement
SCM	Supplementary Cementitious Materials
MARS	Multivariate Adaptive Regression Splines
GGBFS	Ground granulated blast-furnace slag
CNI	Calcium Nitrate Inhibitor
AEA	Air Entraining Agents
HRWR	High-Range Water Reducers
ML	Machin Learning
ANN	Artificial Neural Network
SCC	Self-compacting concrete
PDP	Partial Dependency Plots
ANFIS	Adaptive Neuro-Fuzzy Inference System
MPMR	Minimax Probability Machine Regression
RF	Random Forest
GEP	Gene Expression Programming
DT	Decision Tree
EL	Ensemble Learning
GB	Gradient Boosting
XGB	EXtreme Gradient Boosting
LGBM	Light Gradient Boosting Machine
RFECV	Recursive Feature Elimination with Cross-Validation
MPMR	Minimax Probability Machine Regression
NSGA	Non-dominated Sorting Genetic Algorithm
OA	Optimization Algorithm
PSO	Particle Swarm Optimization
GS	Grid Search
STML	Self-tuning multilayer
TLBO	Teaching-Learning-Based Optimization algorithm
IQR	Interquartile range
MSE	Mean Square Error
MAE	Mean Absolute Error
SHAP	SHapley Additive exPlanation

repair mortars through rapid chloride migration tests to determine chloride diffusivity and correlated these results with practical in-situ indicators such as gas permeability and capillary absorption, demonstrating that simple field tests can serve as reliable predictors of chloride resistance. The RCPT is another method for quantifying chloride migration into concrete, and it has been the most extensively employed technique for this purpose. Introduced by Whiting [12] in 1981, this test evaluated the electrical current passing through a concrete sample. The test's validity is contingent upon a demonstrable correlation being found between the data produced by the RCPT and the corresponding ponding test results.

Although RCPT has been widely employed to evaluate the resistance of concrete to chloride ion penetration, it measures electrical charge instead of actual chloride diffusion [12]. Hence, this measure is sensitive to the presence of certain innovative materials (such as KIM concrete waterproofing admixtures or supplementary cementitious materials). These materials can influence electrical resistivity, and the heat generated during the tests can also influence the outcomes [13]. However, RCPT is useful since it is a practical benchmark for durability evaluation and comparison. While RCPT is classified as a rapid test based on its assumed correlation with the ponding test, it is necessary to allow for a 28-day curing period before carrying out this experiment. Consequently, this test requires considerable time, raw materials, and energy. As a result, researchers worldwide have been developing new prediction models to estimate concrete RCPT before casting the samples [14].

Recently, much research has implemented ML methods to estimate the characteristics of cementitious materials, such as compressive strength, carbonation depth, and punching shear strength of slab-column connections [15]. Similarly, researchers have begun to develop models to predict the rapid chloride permeability or chloride ingress in vibratory-placed concrete. A summary of studies on RCPT prediction of concrete is presented in Table 1. As shown, temperature was not considered the independent variable in most studies. Further, a limited number of independent variables were used in the RCPT prediction models, mostly without detecting the optimal variables that should be used in the modeling to maximize the prediction accuracy. Moreover, the number of data observations was limited to 404 samples. Many studies have shown that increasing the number of data samples

Table 1
Various ML methods for predicting RCPT values.

References	Input parameters	No. Data	Models
Boga et al. [19]	CNI, curing duration and type, GGBFS	54	ANN, ANFIS
Najimi et al. [20]	W/B, C, FA, SF, S, CAg, AEA, HRWR	72	ANN (FF)-ABC
Ghafoori et al. [21]	W/B, C, FA, SF, S, CAg, AEA, HRWR	72	ANN
Kumar et al. [14]	FA, SF, T	360	MARS, MPMR
Mohamed et al. [22]	W/C, C, GGBFS, FA, SF, W, SP, S, CAg, age, charge	86	ANN
Mohamed et al. [23]	W/B, age, C, GGBFS, FA, SF, W, SP, S, CAg	294	ANN
Tran et al. [24]	W/B, exposure condition, C, SF, time exposure, depth of measurement	404	ANN, GB, RF-PSO
Ge et al. [25]	C, W, SF, coarse, S, SP, curing mechanism, exposure time, the exposure condition	216	Hybrid ANN
Tran et al. [26]	FA, W, exposure time GGBFS, SF, C, SP, S, coarse, T, concentration of chloride	386	DT
Kazemi and Gholampour [27]	C, FA, SF, CA, FAg, W, T	360	ANN
Alabdullah et al. [28]	CS, MK, Age, W/B, B, CA, FAg	201	LGBM, XGB
Amin et al. [29]	CS, MK, Age, W/B, B, CA, FAg	201	GEP
Song and Kwon [9]	W/B, C, GGBFS, FA, SF, FAg, CAg, Curing period	120	ANN
Guneyisi et al. [30]	Cement type, W/C, Aggregate/C, SP/C, cure type, Age	90	ANN
Hohod and Ahmed [31]	FA, GGBFS, C, W/B, Age	300	ANN
Khan [32]	C, FA, SF, W, SP, FAg, CAg, Age	N. A.	ANN
Parichatprecha and Nimityongskul [33]	C, FA, SF, W, W/B, SP, CAg, FAg	86	ANN

improves the performance of prediction models (e.g., [16]), and the application of larger datasets can enhance the performance of RCPT prediction models.

Artificial Neural Network (ANN) is the widely used algorithm to predict RCPT. However, powerful ensemble learning techniques have been developed that can outperform ANN in terms of prediction accuracy and computational costs. For example, light gradient boosting machine (LGBM) and extreme gradient boosting (XGB) were found to be more accurate than ANN in prediction studies of other disciplines (e.g., [17,18]). However, their applications to predict concrete RCPT have not received enough attention.

Although machine learning techniques are valuable methods for prediction, a few problems are associated with their use: (1) they have some parameters that should be tuned by users (called hyperparameters), and (2) they are black-box tools. Accordingly, a few methods, like random and grid search have been developed to tune the hyperparameters of machine learning techniques. However, Liashchynskyi et al. showed that those conventional techniques could not find the optimal hyperparameter values and were not computationally efficient [34]. They showed that the application of metaheuristic optimization algorithms was a better approach to tuning the hyperparameters.

Therefore, researchers have developed different metaheuristic optimization algorithms to increase the accuracy of prediction models. For example, Hai et al. adopted three metaheuristic optimization algorithms, particle swarm optimization, marine predator algorithm, and genetic algorithm, to tune hyperparameters [35]. Naseri et al. proposed a multi-objective metaheuristic-based approach to tune hyperparameters using NSGA-III [36]. The outcomes showed that considering multiple metrics to tune hyperparameters could significantly increase the prediction accuracy compared to conventional tuning methods (i.e., Hyperopt, grid search, and random search). Although the proposed model was accurate and efficient, the number of metrics was limited to two.

Although metaheuristic optimization algorithms are efficient methods to tune hyperparameters of machine learning techniques, they have their parameters (called metaheuristic parameters) to be tuned. Hence, using the default metaheuristic parameters questions the validity of the results. On the other hand, tuning the metaheuristic parameters can be a vain loop since they aim to tune hyperparameters. Recently, a parameter-free metaheuristic optimization algorithm has been developed called the Self-tuning multi-layer optimization algorithm (STML) [37]. This algorithm can tune its metaheuristic parameters during the optimization process using an outer layer. Accordingly, STML can be adjusted to tune hyperparameters of machine learning techniques since it can simultaneously tune hyperparameters and metaheuristic parameters.

Regarding the second issue of ML techniques (being black-box), various methods have been developed that can address it, such as sensitivity-based interpretation, SHapley Additive exPlanations (SHAP), and partial dependency plots (PDPs). Sensitivity-based interpretation is the simplest interpretation technique, and it finds the relationship between each independent variable and the dependent variable by changing the value of an independent variable and predicting the dependent variable of the generated data observation. One of the powerful interpretation models is SHAP. This method can present the relative influence of independent variables on the dependent variable (RCPT in this study). Another interpretation method is PDP. PDP can illustrate the non-linear relationship between each single independent variable and the dependent variable [38]. SHAP and PDP are described in the methodology section. These interpretation methods can facilitate a more thorough investigation into the relationship between various variables and the RCPT of concrete.

As seen from the above studies, the optimal feature set (i.e., optimal variables) to maximize RCPT prediction accuracy has rarely been investigated in previous studies. Moreover, a limited number of variables were applied in the modeling, mostly excluding temperature. The number of data observations was limited to 404 samples. Further, applying powerful ensemble learning methods (LGBM and XGB) has not received enough attention in predicting RCPT. Likewise, the application of powerful hyperparameter tuning methods was not considered in previous studies, which can worsen the performance of prediction models. Finally, the relation between optimal variables and RCPT was not investigated using novel interpretation techniques.

To address these limitations, the contributions and objectives of this study are as follows:

- A robust feature selection framework is developed to identify the optimal independent variables applied in the modeling to minimize the prediction error.
- A large-scale dataset is collected and applied to predict RCPT.
- XGB and LGBM are adjusted to predict RCPT, and their performance is compared with the conventional method (i.e., ANN) using multiple metrics to determine the most accurate model.
- A new multi-objective hyperparameter tuning method is introduced that considers three metrics and employs a parameter-free metaheuristic optimization algorithm.
- The influence of considered metrics on the performance of the conventional hyperparameter tuning method (grid search) is examined.
- The non-linear relation between optimal independent variables and RCPT is investigated to identify the optimal range of these variables to minimize RCPT.

2. Methodology

The flowchart depicted in Fig. 1 outlines the process of implementing ML models to predict the RCPT of concrete. As outlined in Fig. 1, this study consists of four consecutive stages: data preparation, tuning and training the models, validating the results and determining the most accurate model, and evaluating the influence of independent variables on RCPT.

2.1. Phase 1. Dataset

To predict the RCPT values, different variables, such as the weight of cement, silica fume, fly ash, binder content, fine aggregate, coarse aggregate, water content, and temperature, were analyzed. Notably, most previous RCPT prediction studies generally neglected the temperature parameter (please see the input parameters in Table 1). To address this gap, we added the temperature as an input parameter to the dataset.

2.2. Phase 2. Feature selection

In this phase, the optimal independent variables were investigated. In other words, all the possible combinations of variables (e.g., material ratios) were generated. Subsequently, a powerful feature selection technique called recursive feature elimination with cross-validation (RFECV) was employed to identify the optimal variables to be applied in prediction modeling. The optimal feature set in this step was used in the following phase.

2.3. Phase 3. Tuning and training the proposed ML models

Training data was used to train all the ML models. We used three different algorithms: the LGBM, XGB, and multilayer perceptron (MLP). These algorithms were tuned using different grid search (GS) models built on R^2, the mean absolute error (MAE), and the mean squared error (MSE), and a new tuning technique was introduced in this study (i.e., STML). Moreover, the ML models were validated by the testing data. Different statistical parameters were utilized to examine the ML models and identify the most accurate ones. The most accurate model was used in the following phase.

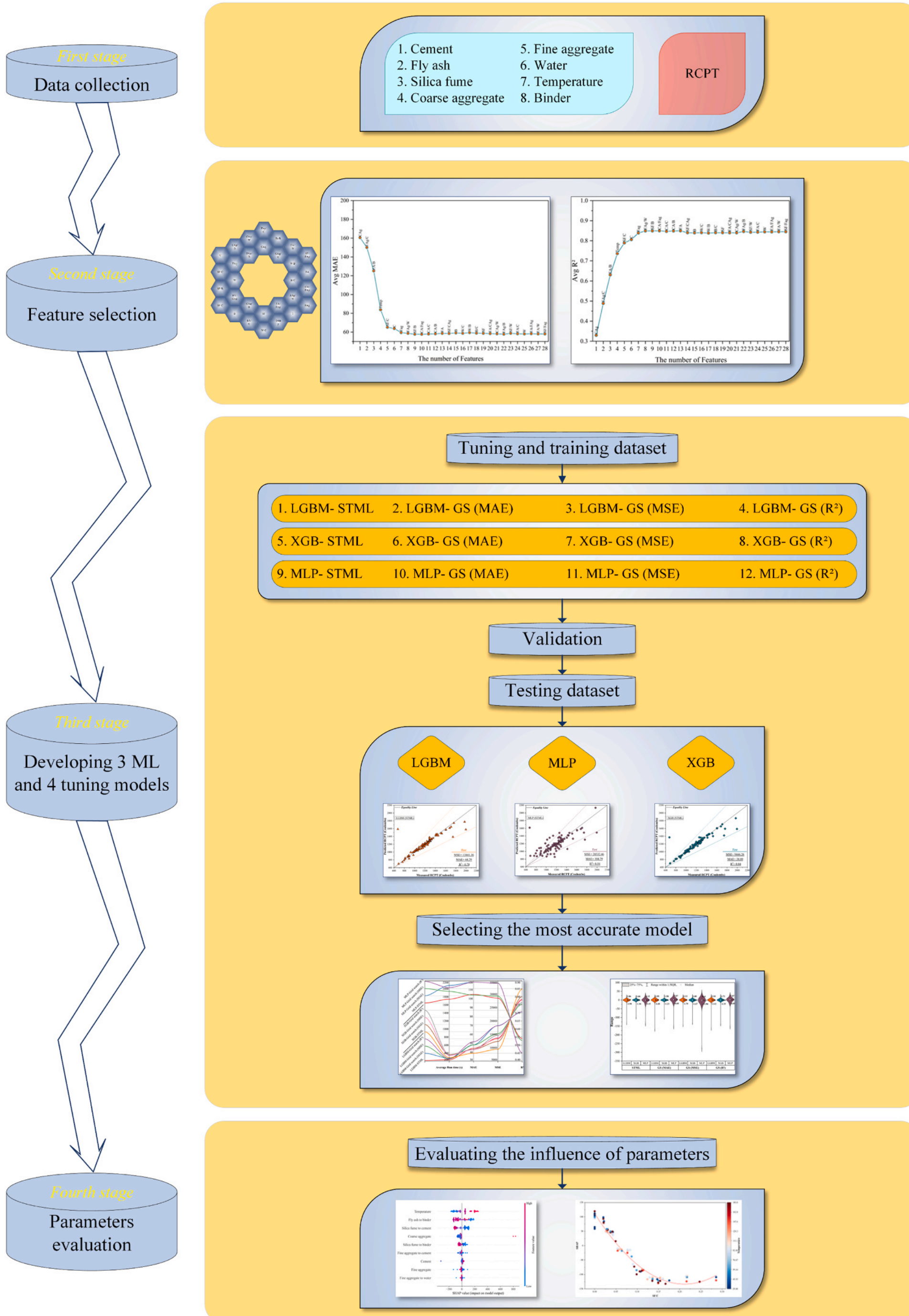

Fig. 1. Flowchart of the methodology.

2.4. Phase 4. Evaluating the relative influence of variables on RCPT

SHAP was employed to determine the relative influence of independent variables on the RCPT of concrete. Furthermore, we perform sensitivity analyses via PDPs to capture the direction of influence of those variables on RCPT at different test temperatures.

2.5. Dataset

The initial data were collected from authentic international publications [39–42]. The initial data included multiple independent variables, including the C (kg/m^3), SF (kg/m^3), FA (kg/m^3), CAg (kg/m^3), FAg (kg/m^3), W (kg/m^3), B (kg/m^3), and T (°C). The dataset consisted of 469 samples. The dataset subsequently underwent a random split into two distinct groups. That is, 80 % of the dataset served as the training dataset, enabling the ML model to be trained. In contrast, the remaining 20 % functioned as the testing dataset, allowing the evaluation of the ML model's prediction performance.

Tables 2 and 3 provide a statistical analysis of the training and testing sets. A visual representation of the variable distributions and their correlation with the outcome (RCPT) can be found in Fig. 2. In this figure, each data is shown by a point, and its frequency distribution is shown on the right axis. In addition, in all graphs, the frequency distribution of RCPT data is shown on the top axis. Fig. 3 shows the Pearson coefficient, which illustrates the correlation between each input and RCPT.

Skw = Skewness, StD = Standard deviation

Table 2
Training data statistical results.

	Unit	Count	Mean	Median	Min	Max	Q_{25}	Q_{75}	StD	Skw
Input										
Portland Cement	kg/m^3	375	340.6	332.3	140	515.1	260.1	430.1	96.5	0.05
Fly Ash	kg/m^3	375	233.2	206.04	0	515.1	103.2	360.5	160.3	0.24
Silica Fume	kg/m^3	375	21.5	20.60	0	51.51	0	41.2	18.6	0.23
Coarse Aggregate	kg/m^3	375	764.09	727.9	617.5	1298.5	672.7	774.6	142.9	1.72
Fine Aggregate	kg/m^3	375	777.09	783.2	536.4	888.3	729.5	829	66.67	−0.71
Water	kg/m^3	375	185.8	185.5	140	225	185.5	185.5	11.61	−0.67
Binder	kg/m^3	375	90.8	90	25	180	30	150	55.13	0.24
Temperature	°C	375	595.5	607.8	350	775.2	551.2	677.3	109.6	−0.65
Output										
RCPT	Coulombs	375	1210.9	1174	633	4641.2	1038	1341.5	345.2	4.69

Table 3
Statistical results for the test data.

	Unit	Count	Mean	Median	Min	Max	Q_{25}	Q_{75}	StD	Skw
Input										
Portland Cement	kg/m^3	94	358.4	361.9	180.3	515.1	288.1	459.7	96.5	−0.07
Fly Ash	kg/m^3	94	226.2	206	0	515.1	58.6	360.6	166.1	0.20
Silica Fume	kg/m^3	94	20.4	10.3	0	51.51	0	41.2	18.6	0.44
Coarse Aggregate	kg/m^3	94	747.08	726.4	617.5	1283.2	672.7	774.6	125.9	2.38
Fine Aggregate	kg/m^3	94	787.2	789.9	533.6	888.3	746.2	835.7	69.9	−0.83
Water	kg/m^3	94	184.9	185.5	140	204.9	185.5	185.5	8.5	−1.75
Binder	kg/m^3	94	605.06	607.8	350	746.9	551.2	677.3	95.4	−0.70
Temperature	°C	94	86.1	90	25	180	30	120	50.88	0.45
Output										
RCPT	Coulombs	94	1194.4	1193.5	673.9	1995	1029	1291	249.3	0.73

Skw = Skewness, StD = Standard deviation

2.6. Feature selection

This study aims to develop a model that estimates the RCPT accurately. To this end, optimal variables should be used in the modeling as independent variables. However, it is unclear which variable set should be used in the model to minimize errors and maximize the performance. Many variables are commonly used to predict different concrete properties. For example, the weights of materials are common independent variables used in modeling [43].

Furthermore, the ratios of the weights of materials are other independent variables that have been applied to develop accurate prediction models. As such, Naseri et al. [44] considered some ratios of material weights as well as the weights of all materials for predicting concrete's CS: W/B, FA/B, SP/B, FAg/Aggregate, and CAg/B. However, in another study, the fly ash-to-cement ratio was considered instead of the fly ash-to-binder ratio [45]. Naderpour et al. [46] used the W/C, water-to-total material ratio, and weight of different aggregates to predict the concrete's CS.

Although material weights and some of their weight ratios have been extensively applied to predict different properties of concrete, the ratio of the weights of materials that should be considered independent variables is unclear. Many ratios can be generated by dividing the weights of all the materials; however, there is a lack of information about the optimal variables, leading to maximizing the prediction accuracy. Considering all the possible weight ratios and all the material weights simultaneously increases the number of variables, which may lead to many analytical problems, such as accuracy reduction, overfitting, and increased computational costs [47]. To address this, applying a powerful feature selection technique is essential [48].

To detect the optimal variables, all the possible variables (28 variables) were considered in the initial dataset, including the C, W, CAg, FAg, FA, SF, B, T, and the B/C, SF/C, FA/C, W/C, CAg/C, FAg/C, SF/B, FA/B, W/B, CAg/B, FAg/B, SF/W, FA/W, CAg/W, FAg/W, SF/FAg, FA/FAg, CAg/FAg, SF/CAg, and FA/CAg. Then, RFECV was used to detect the optimal variables to be applied in the prediction modeling. The RFECV is a wrapper feature selection that identifies the most relevant features explaining the output variable [49]. The number of folds in RFECV was set to five, and we applied RFECV rather than the conventional RFE as the feature selection method to avoid overfitting. The RCPT of the concrete was considered the dependent variable (output). The average MAE and average R^2 were considered metrics to determine the optimal features. To this end, RFCV was run 27 times for the number of features from 1 to 27. The selected features of the models were subsequently applied to run new models, and their MAEs and R^2 values were compared to identify the optimal feature set. Accordingly, the optimal feature set that leads to the prediction of the RCPT with the highest possible accuracy was applied in further analyses.

2.7. Prediction methods

2.7.1. Machine learning methods

Three ML methods were used for modeling: LGBM, XGB, and MLP. LGBM and XGB were selected since they showed superior performance compared to other machine learning techniques in terms of prediction accuracy in previous studies (discussed in the following paragraphs). Also, MLP is chosen since ANN is the conventional and well-known method for predicting the durability of concrete (e.g., [9,19,20,22,23,27,30]), and MLP is an accurate and complex ANN. These methods are described in this section, respectively.

LGBM is a novel open-source ML technique initially introduced by Microsoft. LGBM is a new update of the conventional GB method, and it was developed to address the major deficiencies of that method. LGBM generates a particular number of weak learners (i.e., decision trees). These weak learners are combined to form a powerful ensemble learning model [50]. This method applies two innovative approaches to train each learner using a small portion of the entire dataset and to handle high-dimensional sparse features more efficiently. LGBM is famous for parallel learning and memory usage reduction, and as a result, it is an efficient and quick method for complex problem modeling [51]. LGBM structure is depicted in Fig. 4.

LGBM is applied as one of the learning algorithms since it has surpassed many statistical analyses and learning algorithms when

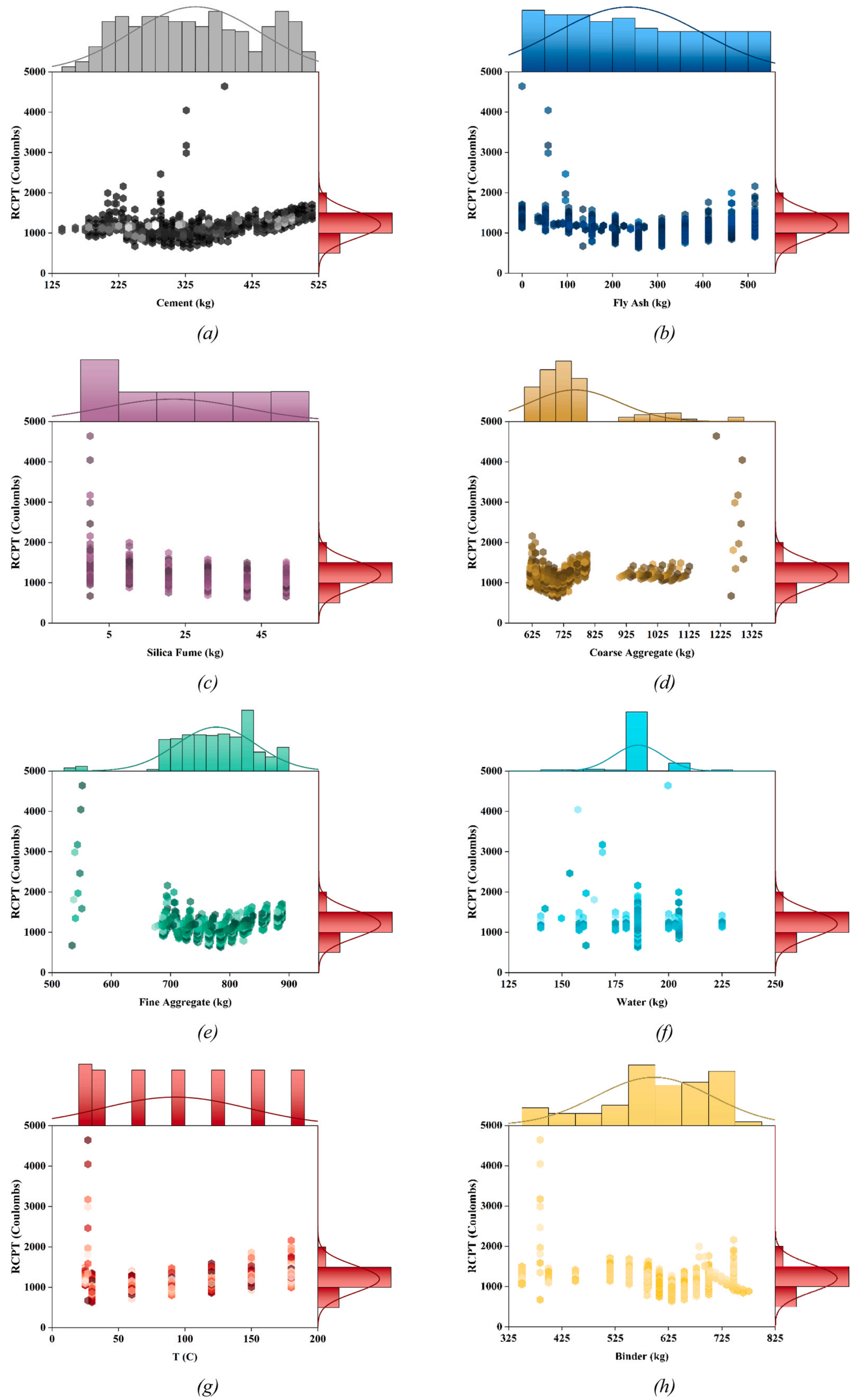

Fig. 2. Frequency distribution and correlations for input and output variables.

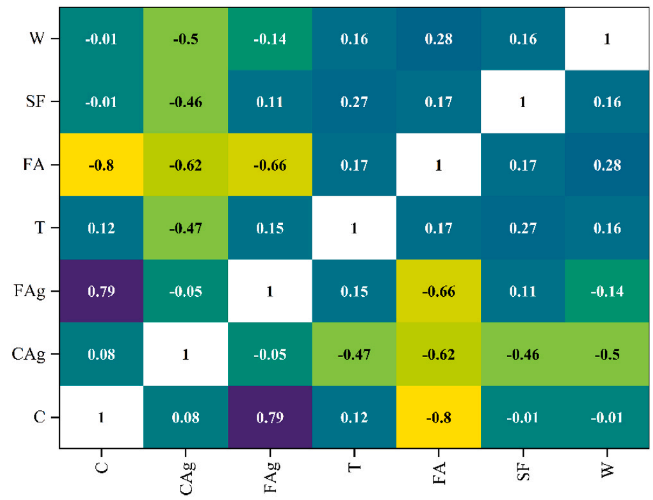

Fig. 3. Pearson coefficient of variables.

comparing prediction accuracy and computational cost, e.g., recurrent neural networks, back-propagation neural networks, stochastic gradient boosting [18], support vector machines [52], multiple linear regression [53], k-nearest neighbors [54], random forests [55], multi-layered perceptron [56], and categorical boosting [57]. Therefore, it can be a candidate for predicting RCPT with the highest accuracy.

XGB is a strong learning model for various prediction tasks. This method optimizes the implementation of the gradient boosting method. Like LGBM, XGB is an EL method that creates different weak learners and merges them to create a powerful ensemble model [58]. XGB applies optimization modeling that increases the model accuracy during the iterations. In short, the internal optimization algorithm leverages first and second-order gradients to refine the weak learners within the final ensemble structure [59]. To prevent overfitting, a regularization term is typically added to the optimization problem's objective function. Fig. 5 illustrates the structure of the XGB model. XGB could also be a potential method to boost the performance of RCPT prediction models since it has consistently outperformed many statistical learning tools in terms of accuracy, such as deep neural networks, random forests [17], adaptive boosting [50], logistic regression [60], categorical boosting [61], support vector machines [62], multinomial logit, and mixed multinomial

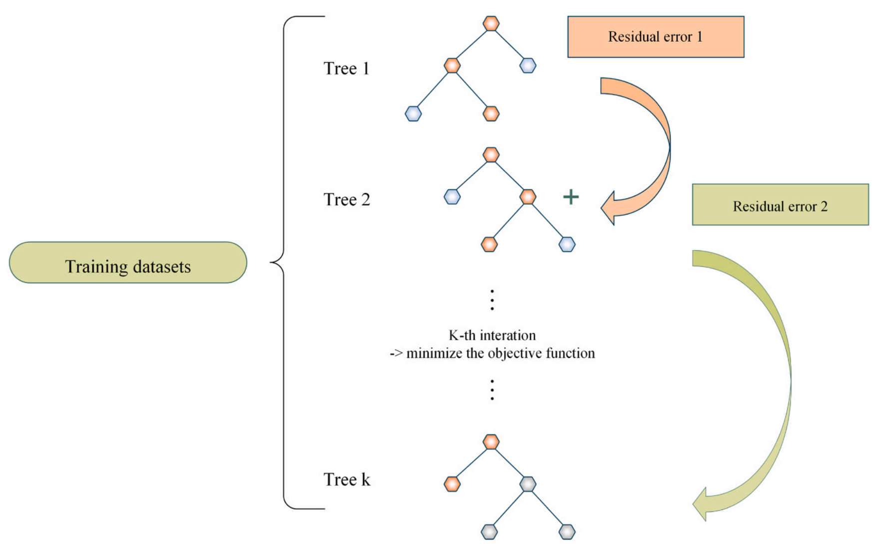

Fig. 4. Structure diagram of LGBM.

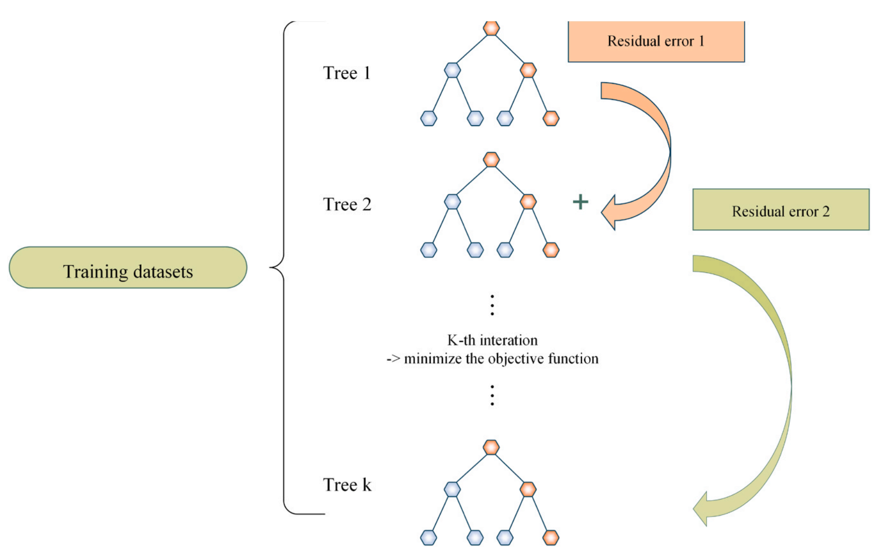

Fig. 5. Structure diagram of XGB.

logit models [63].

Neural networks are the conventional and well-known methods for predicting the durability of concrete (e.g., [9,19,20,22,23,27,30]). Therefore, a deep neural network model, called MLP, is used in this investigation to evaluate the effectiveness of other prediction models. Like other neural networks, the MLP contains three layer types: input, hidden, and output. The MLP is similar to ANNs but comprises more than one hidden layer [47]. Each hidden layer includes some processing units known as neurons. Various connection weights are applied to unidirectionally connected neurons. The input layer collects the data and prepares it for modeling. The adjusted data are subsequently transferred to the first hidden layer, and weights are assigned to each neuron. Then, the neuron connection information and its weights are sent to the following hidden layer. The role of the output layer is to transform the information of the final hidden layer and to present the predicted value [64]. The MLP structure is demonstrated in Fig. 6.

2.7.2. *Hyperparameter tuning*

This study introduces a new hyperparameter tuning method via a recent metaheuristic optimization algorithm. The developed method's results were compared with those of GS, a conventional tuning method. GS was used for this comparison since it is almost a parameter-free tuning method. However, other tuning methods (e.g., Bayesian optimization or Optuna) generally require some parameters to be tuned, such as the number of trials or kernel type. In this section, the developed method is first presented, and then GS is described.

The STML is a recently introduced parameter-less metaheuristic algorithm that can automatically tune its parameters (self-tuning). Hence, this algorithm does not require manual fine-tuning. This algorithm includes two optimization layers: inner and outer layers. The inner layer acts as a stochastic population-based method and aims to optimize the objective function by finding the optimal values of the dependent variables. TLBO is generally used to optimize the decision variables in the inner layer. The inner layer includes two phases: exploitation and exploration. The exploitation is responsible for local search, and the exploration prevents the model from being confined to local optima and helps the model investigate new search spaces. In the exploitation phase (called the teacher phase), each solution vector (in this study, a set of hyperparameters) moves toward the best solution vector of the iteration (called the teacher) to find better solutions nearby. Further, each solution vector (called a learner in TLBO) randomly interacts with other solution vectors, and solution vectors with lower fitness values push toward learners with higher fitness values. This process avoids the problem of local minima [65]. The pattern search algorithm is utilized to dynamically optimize the decision variables of the outer layer (metaheuristic parameters).

This study synchronizes STML with machine learning techniques to tune their hyperparameters. The proposed model can be mathematically formulated as follows:

$$Z_1 = \text{maximize}\text{R}^2{}_{(\text{K-CV})} \tag{1}$$

$$Z_2 = \text{minimize}\text{MAE}_{(\text{K-CV})} \tag{2}$$

$$Z_3 = \text{minimize}\text{MSE}_{(\text{K-CV})} \tag{3}$$

s.t:

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(y_k - \tilde{y}_k)^2}{\sum_{k=1}^{n}(y_k - \bar{y})^2} \tag{4}$$

$$MAE = \frac{\sum_{k=1}^{n}|y_k - \tilde{y}_k|}{n} \tag{5}$$

$$MSE = \frac{\sum_{k=1}^{n}(y_k - \tilde{y}_k)^2}{n} \tag{6}$$

$$HP_i^D \in Set_i \forall i \in I \tag{7}$$

$$HP_j^C \geq HP_j^{\min} \forall j \in J \tag{8}$$

$$HP_j^C \leq HP_j^{\max} \forall j \in J \tag{9}$$

Where $R^2_{(K\text{-}CV)}$, $MAE_{(K\text{-}CV)}$, and $MSE_{(K\text{-}CV)}$ are the average R^2, average MAE, and average MSE over K folds, respectively. Furthermore, $\boldsymbol{y_k}, \boldsymbol{\tilde{y}_k}, \boldsymbol{\bar{y}}$, and n denote the actual value of sample k, the predicted value of sample k, the mean of actual values, and the number of data samples, respectively. HP_i^D, I, HP_j^c, and j are the values of the discrete-ranged hyperparameter i, the number of discrete-ranged hyperparameters, the value of the continuous-ranged hyperparameter j, and the number of continuous-ranged hyperparameters, respectively. HP_j^{max} and HP_j^{min} show the highest and lowest defined values for the continuous-ranged hyperparameter j.

In the introduced optimization framework, Eqs. (1), (2), and (3) are the objective functions, and the model aims to maximize R^2 while minimizing the MAE and MSE. In the optimization model, k-fold cross-validation is used to optimize the metrics, considering $\text{k} = 5$. Eqs. (4) to (9) are the constraints of the optimization problem. Eqs. (4) to (6) are used to evaluate the metrics. Eq. (7) ensures that valid values are assigned to discrete-ranged hyperparameters. Eqs. (8) and (9) guarantee that only values in the given (allowed) ranges are allocated to continuous-ranged hyperparameters in the optimization process. The proposed model aims to optimize the hyperparameters of ML techniques via a multi-objective parameter-free optimization approach (multi-objective STML).

A schematic diagram of the proposed hyperparameter tuning method is shown in Fig. 7. As indicated, this method includes three parts:

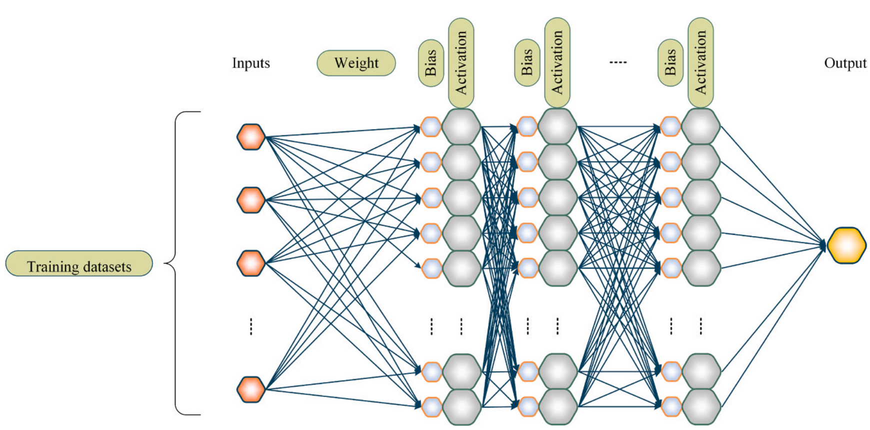

Fig. 6. Structure diagram of MLP.

- Outer optimizer: This part aims to optimize the parameters of the metaheuristic algorithm.
- Inner optimizer: This part receives the metaheuristic parameters (TLBO algorithm) from the outer optimizer and sends the convergence information to that layer. The inner optimizer sends hyperparameter values to the machine learning algorithm and receives the corresponding hyperparameter values' performance indicators (i.e., R^2, MAE, and MSE).
- Machine learning: This part receives different hyperparameter sets from the inner optimizer, runs the model, and calculates the performance indicators for each hyperparameter set. Then, the performance indicator values are sent to the inner optimizer.

GS is a comprehensive search method for tuning hyperparameters. GS is a widely used method that predefines a given number of values for each hyperparameter. All the possible combinations of the predefined values are considered potential solutions. The machine learning model is subsequently run for each combination as the hyperparameter. The performance of the models is evaluated via a single performance metric (e.g., mean absolute error). In this study, three GS models are developed, each of which uses a different metric: MSE, MAE, and R^2. This analysis can investigate the impact of the performance metric on the GS's performance.

2.7.3. Statistical criteria

Evaluating the performance of the constructed predictive model is a necessary step [66]. Consequently, in this study, the performance of the proposed model was assessed via three statistical parameters: MSE, MAE, and R^2. More accurate forecasting models exhibit lower MSE and MAE values [67]. Indeed, R^2 measures the discrepancies between the actual and predicted values, ranging from 0 to 1. A higher R^2 indicates a better-proposed model [68]. The expressions for the three statistical indices are as follows [69]:

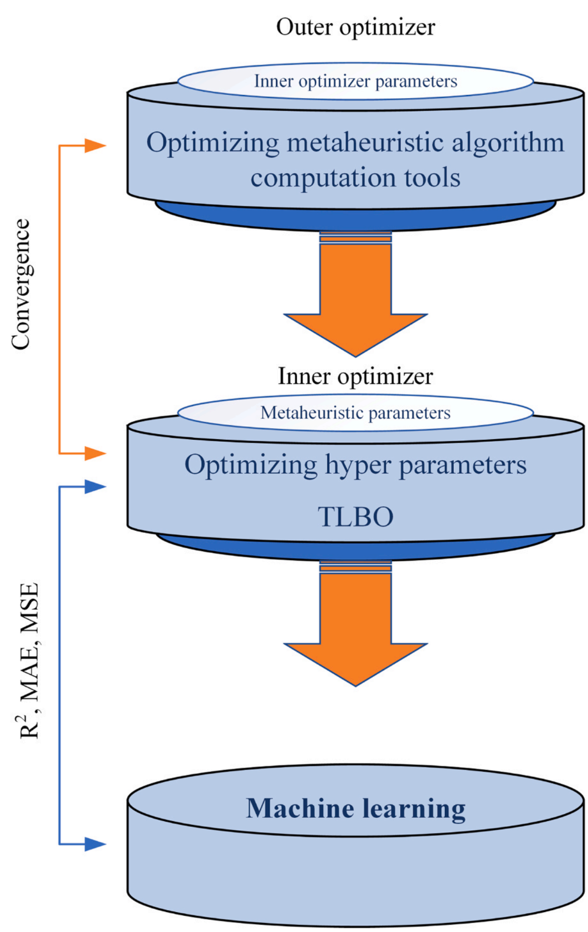

Fig. 7. The schematic flowchart of the introduced tuning method.

$$\boldsymbol{MSE} = \frac{1}{n}\sum_{i=1}^{n}(t_{0,j} - o_{t,i})^2 \tag{10}$$

$$\boldsymbol{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|t_{0,j} - o_{t,i}\right| \tag{11}$$

$$\boldsymbol{R}^2 = \frac{\left(n\sum t_{0,j}o_{t,i} - \sum t_{0,j}\sum o_{t,i}\right)^2}{\left(n\sum {t_{0,j}}^2 - \left(\sum t_{0,j}\right)^2\right)\left(n\sum {o_{t,i}}^2 - \left(\sum o_{t,i}\right)^2\right)} \tag{12}$$

where, $t_{0,\ j}$ signifies the actual RCPT value of the j^{th} sample, $o_{t,\ i}$ corresponds to the RCPT value of the i^{th} sample predicted by the ML model, and n is the number of samples.

2.8. Interpretation techniques

Since most ML techniques are black-box methods, interpretation techniques are vital for capturing the impact of independent variables on RCPT. Two interpretation methods are applied, as described in this section, to address this.

2.8.1. SHAP analysis

To better explain the prediction and relative significance of each input, researchers have employed a game theory technique called SHAP. This approach could quantify each input's contribution to the analysis [70]. The SHAP value represents the average influence of each input variable on RCPT. The absolute value of SHAP is used to calculate the variables' importance. Variables with higher absolute SHAP values are considered significant variables [71]. A positive SHAP value indicates that the independent variable increases the dependent variable. A negative SHAP value indicates that the independent variable decreases it. The magnitude shows the strength of that influence (called relative influence). Using SHAP values, the variables with the highest relative influence of independent variables on RCPT are identified. Also, the general relationship between the independent variables and RCPT is captured. However, SHAP values can not be applied to determine the optimal values of independent variables leading to minimum RCPT values. To address this, dependence plots are employed.

2.8.2. SHAP dependence plot

To visualize the output of ML models, Friedman introduced the widely used technique called PDPs [72]. By analyzing the PDP results, it is possible to detect the relation between the input and output variables, whether they are linear, monotonic, or complex. The PDP can show how the predicted value changes based on different values of independent variables. To measure the average partial dependence function, $\boldsymbol{F_s}$, on a subset of predictors $\boldsymbol{K_s}$, the following method is used [73]:

$$\overline{F_s}(K_s) = \int \widehat{F}(K_s, K_c)dp(K_c) \tag{13}$$

The input variables have a target feature denoted as $\boldsymbol{K_s}$. Assuming that $\boldsymbol{K_c} \cup \boldsymbol{K_s}$ is equal to S when $\boldsymbol{K_c}$ serves as the complement, the marginal effect of $\boldsymbol{K_c}$ can be represented by $\boldsymbol{dp}(\boldsymbol{K_c})$. The average partial dependence function for data $\{S_i, i - 1,\ 2,\ \ldots, n\}$ obtained via Eq. (13) can be constructed as follows:

$$\overline{F_s}(z_s) = \frac{1}{n}\sum_{i=1}^{n}\overline{F}(K_s, K_{i,c}) \tag{14}$$

Where n represents the total number of samples. In the training dataset, $\boldsymbol{K_{i,c}}$ corresponds to the actual value of the i^{th} variable. In this study, a more comprehensive analysis was conducted by utilizing a single partial dependence plot (PDP-1D), as described by Goldstein et al. [74]. It represents how a specific input feature impacts the predicted output. In this study, the SHAP dependency plot is used to illustrate PDPs. The results of SHAP and PDP depend on the accuracy of the performance of the base machine learning technique, and reducing the accuracy may lead to inaccurate outcomes in the interpretation techniques. Therefore, we first compared the performance of the prediction model and identified the most accurate one. Subsequently, the best-performing model was synchronized with interpretation techniques, according to details provided by Naseri et al. [38].

3. Results

3.1. Optimal feature set

We used feature selection to identify the optimal set and number of variables for predicting RCPT. The outcomes of the feature selection analysis are illustrated in Fig. 8. As shown, the optimal number of features to predict RCPT is nine, representing the highest average R^2 and the lowest average MAE. The independent variables in the optimal feature set included the coarse aggregate (CAg), fine aggregate to cement ratio (FAg/C), fly ash to binder (FA/B), temperature (T), silica fume to cement (SF/C), cement (C), fine aggregate (FAg), fine aggregate to water (FAg/W), and silica fume to binder (SF/B). These variables were used in further analyses. The optimal feature set found here is

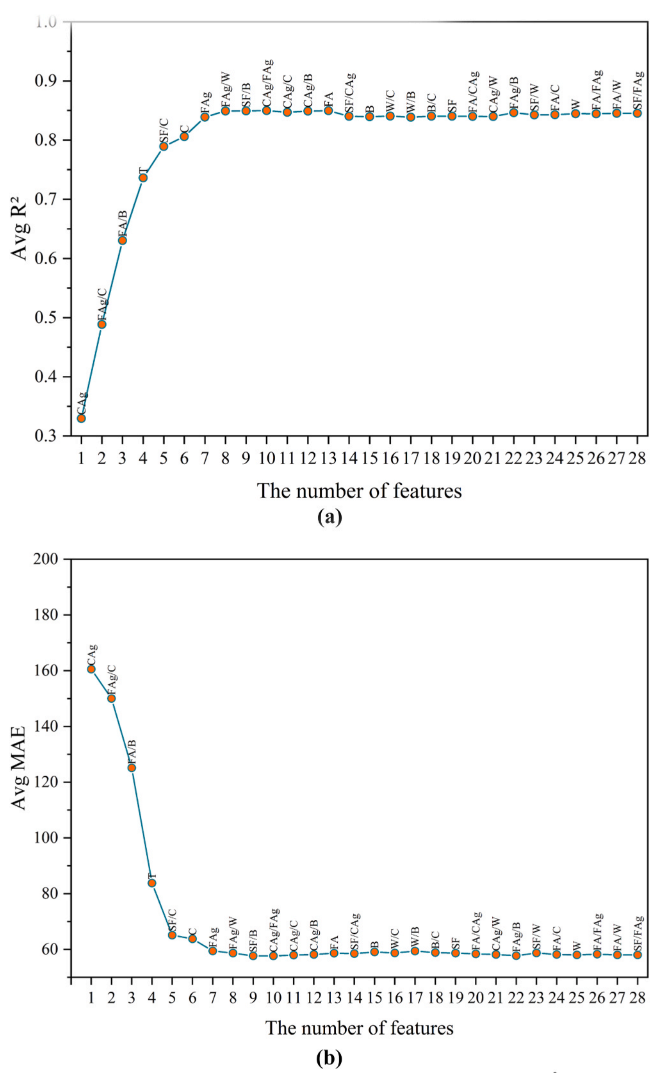

Fig. 8. Feature selection of various possible inputs based on a) R^2 and b) MAE.

applied to train and tune machine learning algorithms.

Selecting the most appropriate hyperparameter tuning technique is often vital to achieve satisfactory performance while minimizing the computational cost. This can be attributed to the numerous hyperparameter optimization techniques mentioned in the literature. However, the application and development of powerful tuning techniques have yet to receive attention in studies predicting the durability of concrete. As another research gap, the influence of selected metrics on the performance of GS (as the most commonly used tuning technique) has not yet been investigated. Hence, four hyperparameter optimization techniques are evaluated to detect the most effective technique. The hyperparameters of XGB, LGBM, and MLP models are individually fine-tuned in their search spaces. Table 4 shows the tuned hyperparameters for the developed models.

3.2. Performance of the developed models

The performances of ML models on validation and testing sets are presented in Table 5. As can be seen, the values shown in this Table are approximately equal to (or even less than) R^2 and MAE of the feature selection method, shown in Fig. 8. Therefore, it can be postulated that the RFECV did not lead to overfitting since the testing data of optimized models obtained less MAE and higher R^2 than those of the average cross-validation of RFECV. Fig. 9 also demonstrates the models' performance on the testing data. Notably, both the XGB and LGBM models outperform the MLP model (the conventional technique) in predicting RCPT. The XGB and LGBM models tuned by a grid search underperformed the models that were STML optimized. In other words, this study's developed hyperparameter tuning model (STML) significantly outperforms different types of GS in terms of prediction accuracy.

Applying MSE as the metric in the GS models generally led to better performance than other metrics. For example, in the MLP, GS (MSE) outperforms GS (MAE) and GS (R^2) in terms of all performance indicators. Similarly, in the LGBM, the best GS model is GS (MSE). In XGB, the GS (MSE) reached the minimum MAE compared with the other grid search models, whereas the GS (MAE) was the best GS model based on the MSE and R^2. Therefore, it is recommended that the MSE be considered the metric when using GS for RCPT prediction.

As shown in Fig. 9, XGB-STML is the most accurate model for predicting RCPT. XGB-STML has the highest testing data R^2 and the minimum testing data MSE and MAE. The performance of XGB-GS (MAE), XGB-GS (MSE), and LGBM-STML is also acceptable. Therefore, XGB-STML, the best-performing model, is applied in further analyses.

Figs. 10, 11, and 12 visually represent testing data error histograms of different machine learning techniques tuned by STML. These models are the best versions of XGB, LGBM, and MLP. The outcomes demonstrate satisfactory concordance between the estimated and actual values in the testing data. The vicinity of the data points to the equality line suggests that the models tuned by STML have satisfactory accuracy in predicting the RCPT. These figures present the data and highlight the error range, approximately 20 %, as the dashed line shows. On these figures, we considered the error range of ±20 % as an acceptable error range, since this level is a common practice in evaluating the performance of machine learning models in predicting various characteristics of concrete [75]. According to ASTM C1202–12, the variation of two RCPT tests on concrete samples from the same batch and of the same diameter by the same operator should be less than 42 %. Therefore, the acceptable variation in the two replications of RCPT tests is 42 %, while the developed methods in this study can predict the RCPT of samples with an error of less than 20 % for all unseen data (except one sample), which is much better than the acceptable level. Accordingly, the developed model in this study can be applied to predict the RCPT concrete in real-life applications before producing the samples. This process can save materials, reduce energy use, reduce the GHG emissions produced in the concrete industry, and reduce waste generation. The high R^2 value and low MSE and MAE values show that the STML model is

Table 4
Hyperparameter values.

Method	Hyperparameter	Description	STML	GS (R^2)	GS (MAE)	GS (MSE)
XGB	Number of estimators	Number of weak learners	234	450	450	450
	Learning rate	Step size	0.115	0.15	0.05	0.2
	Maximum depth	The maximum depth of trees	3	5	5	5
	Minimum child weight	The minimum sum of instance weights in a node	3.998	10	1	8
	Gamma	The minimum reduction in loss needed to make a split	0.942	4	3	4
LGBM	Number of estimators	Number of weak learners	302	250	500	500
	Learning rate	Step size	0.3	0.1	0.1	0.07
	Maximum depth	The maximum depth of trees	5	4	7	4
	Number of leaves	The maximum number of leaves per tree	66	30	30	30
	Min child samples	The minimum number of data samples required in a leaf	11	10	20	10
MLP	First hidden layer size	Number of neurons	96	32	32	64
	Second hidden layer size	Number of neurons	110	128	128	128
	L2 Regularization	A penalty to the loss function	0.079	0.001	0.001	0.001
	Learning rate	Step size	0.023	0.005	0.005	0.005

Table 5
Statistical details of the developed models.

Models	Optimization Methods	Validation			Test		
		MAE	MSE	R^2	MAE	MSE	R^2
LGBM	STML	53.442	17,963.69	0.867	44.291	13,441.37	0.783
	Grid search (MAE)	53.014	18,809.88	0.838	48.257	18,252.03	0.706
	Grid search (MSE)	56.542	17,373.35	0.864	48.104	15,248.33	0.754
	Grid search (R^2)	56.086	17,633.7	0.868	51.717	18,201.34	0.707
XGB	STML	48.296	21,520.06	0.854	38.893	9446.269	0.848
	Grid search (MAE)	40.875	23,769.3	0.841	43.074	12,345.85	0.801
	Grid search (MSE)	49.904	19,035.37	0.886	42.801	13,751.32	0.778
	Grid search (R^2)	51.042	17,881.03	0.855	43.281	14,654.77	0.764
MLP	STML	108.547	25,290.6	0.764	104.798	28,532.47	0.540
	Grid search (MAE)	110.606	27,029.25	0.761	118.417	34,103.03	0.451
	Grid search (MSE)	110.678	26,224.07	0.760	103.962	29,272.69	0.529
	Grid search (R^2)	110.606	27,029.25	0.761	118.417	34,103.03	0.451

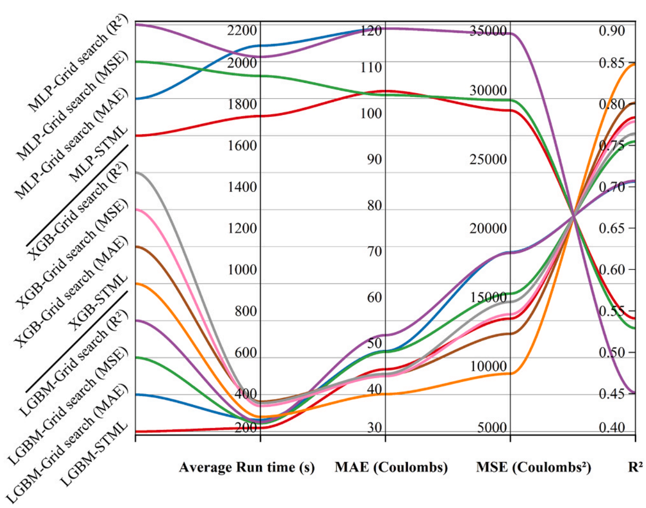

Fig. 9. Comparison between developed models based on various factors.

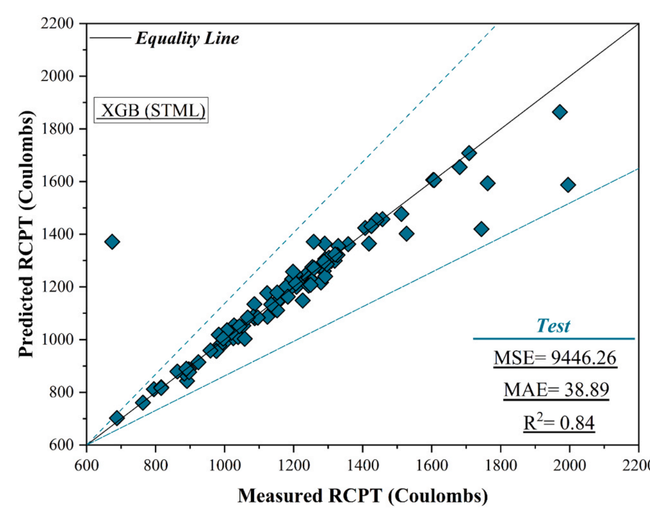

Fig. 10. Error of the XGB-STML model for testing data.

highly precise in RCPT prediction.

The error distribution of the predicted RCPT values is shown in Fig. 13 for the three ML algorithms (LGBM, XGB, and MLP) tuned with four different techniques (STML, GS (MAE), GS (MSE), and GS (R^2)). A violin plot was employed to display the error distribution and density. Regarding accuracy, the XGB-STML model outperforms the best by exhibiting a smaller error range and a density of near zero.

By analyzing the data presented in this section, it can be inferred that the STML outperforms GS models. Additionally, the XGB-STML model achieves the highest level of performance among all tuned models, and it is used to investigate the influence of optimal features on RCPT. However, the STML is only compared with GS in the current study, since both are almost parameter-free, and GS has been the most used method in the literature. Other tuning methods, such as Optuna and Hyperopt, have often been used in previous studies. The advantage of STML over these techniques is that it is a parameter-free algorithm and does not need to be tuned. However, Optuna and Hyperopt need to be tuned themselves, and then, they can be applied to tune the hyperparameters of prediction techniques. Hence, applying STML is recommended when tuning Optuna and Hyperopt is complicated. Moreover, Optuna and Hyperopt can only use a single metric to tune the hyperparameters (like GS), but STML can apply multiple performance indicators

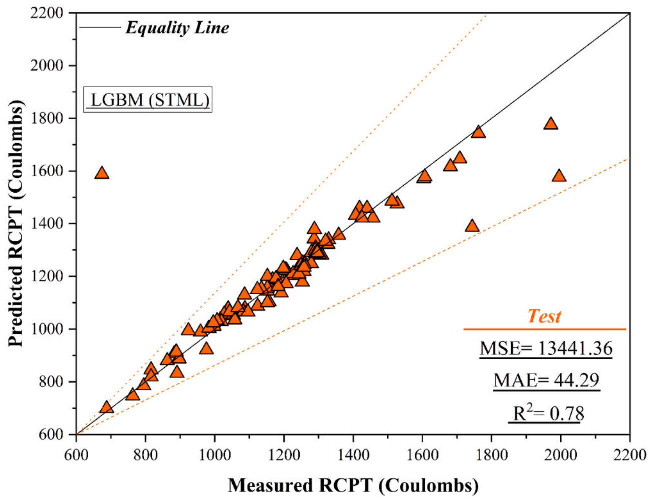

Fig. 11. Error of the LGBM-STML model for testing data.

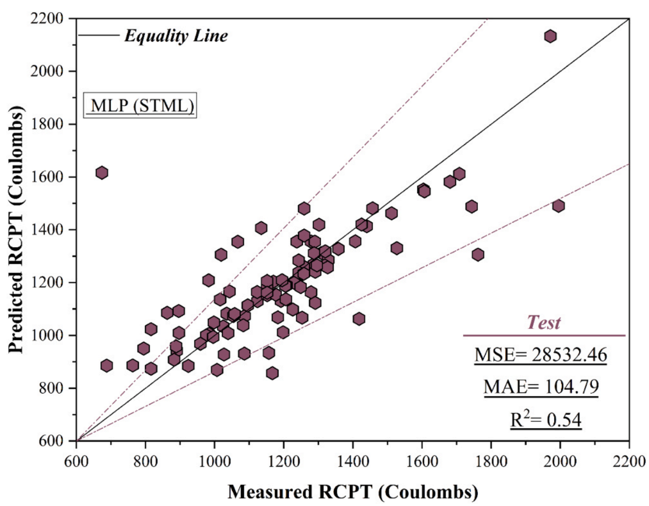

Fig. 12. Error of the MLP-STML model for testing data.

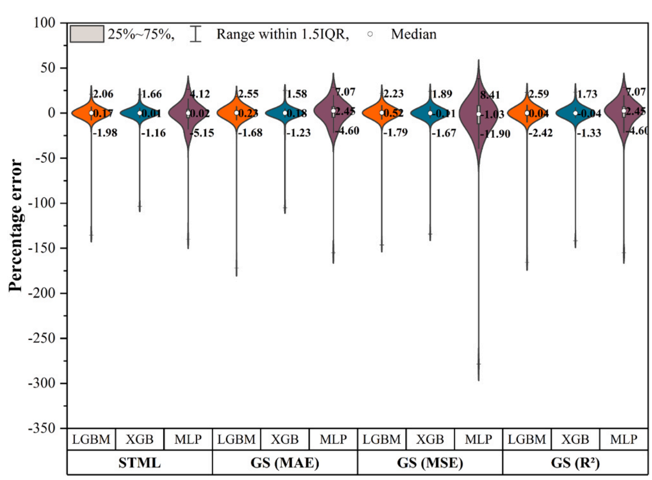

Fig. 13. Error Plot of LGBM, XGB, and MLP.

simultaneously, improving the models' performance. We compared STML with Optuna and Hyperopt on a few scenarios, and STML outperformed them in performance metrics (e.g., MAE and MSE). However, Optuna and Hyperopt could converge faster. Therefore, STML performs better when the objective is to maximize the prediction performance, while Optuna and Hyperopt should be used when the primary aim is to minimize computational costs. We did not present the results of Optuna and Hyperopt in this paper for two reasons: 1. The scope of this study is to analyze parameter-free tuning techniques; 2. The length of the paper would be more than a standard level.

To compare the performance of the developed model, we compared the testing data MAE of the best-performing model (XGB-STML) with the previous studies. To this end, we collected the MAE of the previous studies that predicted RCPT and reported the models' performance on testing data. Moreover, in the cases where more than one machine learning methods were taken into account, we considered the most accurate model of each study (i.e., minimum MAE). The results of the comparison are presented in Fig. 14. As shown, the developed model in this study can significantly reduce the average error by applying the optimal features, considering temperature as an independent variable, and tuning the hyperparameters with a novel approach.

3.3. Influence of optimal features on RCPT

3.3.1. Feature importance analysis

Fig. 15 illustrates the relative influence of each optimal variable on RCPT. The figure clearly shows that RCPT is influenced primarily by temperature. Despite its importance, this variable has rarely been utilized in previous studies, as noted. The second crucial factor is the fly ash to binder ratio. The proportion of silica fume to cement is the third significant input variable. The figure also reveals that the importance of the fine aggregate to water ratio is relatively low, whereas the content of fine aggregates has greater significance. Fig. 15 illustrates the relation between the optimal variables and RCPT, e.g., increasing the temperature corresponds to an increase in the RCPT value. However, in some cases, detecting the relation is not clear. Hence, PDP is used to better understand such relations, and its results are presented in the following part.

3.3.2. Partial dependence plot

PDP was performed to indicate how the input variables affect the RCPT values of concrete, and the results are shown in Fig. 16. PDP has the advantage of providing a straightforward way to determine the nonlinear relation of each independent and dependent variable (here, RCPT). The PDP analysis assessed each input variable's effect while keeping all other variables unchanged. As the temperature is the most critical variable for predicting RCPT, the interaction between temperature and the other variables was analyzed in the PDP analyses.

Increasing the fly ash to binder ratio from 0 to approximately 0.45 reduces the RCPT, indicating a lower permeability level and improving the concrete's durability. This is evident from the steady linear decrease in the partial dependence, as illustrated in Fig. 16a. Increasing the ratio

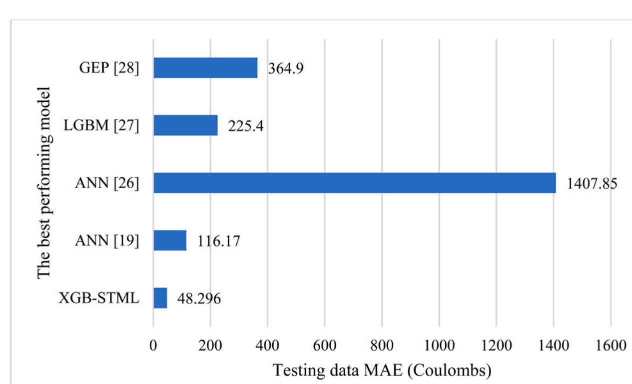

Fig. 14. Comparing the XGB-STML with the previous RCPT prediction models.

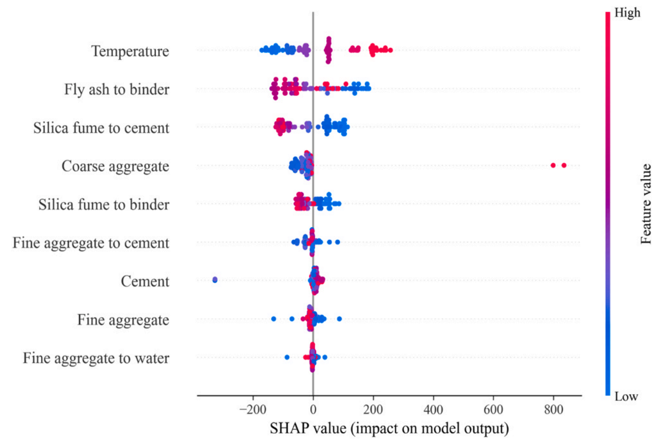

Fig. 15. Feature importance analysis using the SHAP.

of fly ash to binder from 0.45 to 0.7 appears to increase the RCPT and decrease the durability. Therefore, the optimal fly ash to binder ratio was between 0.35 and 0.6.

Dhanya and Santhanam investigated the influence of fly ash on RCPT, and the results suggested that increasing the content of fly ash (i.e., the percentage of replacement with cement) reduces the RCPT [76]. Arunchaitanya and Dey showed that raising the fly ash-to-cement ratio to a threshold (0.333) reduces the RCPT, and after that, RCPT starts to increase [77]. Therefore, their results support the outcomes of this analysis. As a pozzolanic substance, fly ash contains SiO_2 and Al_2O_3. These materials are combined with calcium hydroxide, which is generated during cement hydration. The mentioned reaction generates calcium aluminate hydrate (CAH) and calcium silicate hydrate (C-S-H). CSH and CAH lower the permeability and porosity and increase density by filling the pores in the concrete matrix [39]. Hence, low-content fly ash improves the durability of concrete (i.e., reduces the RCPT). However, high-content fly ash replacement leads to insufficient cement and, as a result, slower cement hydration and more voids between unreacted fly ash particles [78].

Fig. 16b demonstrates a negative relationship between RCPT and the silica fume-to-cement ratio. That is, by increasing the silica fume to cement ratio from 0 to 0.15, the RCPT linearly decreases. Following this, the curve levels off, indicating that raising the silica fume-to-cement ratio from 0.15 to 0.3 has no significant effect on RCPT. Therefore, ratios of 0.15 and 0.3 are optimal for achieving maximum durability. These findings align with earlier research, such as the study by Bagheri et al., which demonstrated that increasing silica fume content in concrete mixtures lowers RCPT and enhances durability [79]. Similarly, Chaudhary and Sinha showed a negative relation between silica fume content and RCPT of concrete [80]. Like fly ash, silica fume reacts with calcium hydroxide and generates C-S-H gel. The fine particles of silica fume and C-S-H gel fill the pores, which improves the microstructure of the concrete, and hence, RCPT is reduced [81].

Fig. 16c shows the relation between the coarse aggregate and its impact on the RCPT. Although the relationship is not linear, when the coarse aggregate content reaches approximately 700 kg/m^3, the RCPT is minimal, and a reduction or increase in coarse aggregate content can increase the RCPT. Therefore, the optimal coarse aggregate content to minimize RCPT is approximately 700 kg/m^3.

When the fine aggregate-to-cement ratio is increased to approximately 2, there is a consistent linear decrease in partial dependence. This implies that increasing the fine aggregate-to-cement ratio within this range reduces the RCPT, as shown in Fig. 15d. Conversely, an increase in the fine aggregate to cement does not impact the RCPT, regardless of the amount of aggregate present. The minimum RCPT is reached when the fine aggregate-to-cement ratio is between 2 and 2.5.

As shown in Fig. 16e, RCPT and fine aggregate content show an inverse relationship. Increasing the content of fine aggregates reduces the RCPT and improves its durability. Fine aggregates fill the small pores, and it can be the reason why increasing fine aggregate contents improves durability [82]. According to Fig. 15f, the RCPT values remain unaffected by the cement content. However, increasing the cement content to greater than 450 kg/m^3 can reduce the RCPT.

Increasing the fine aggregate to water ratio from 3.3 to approximately 4.2 reduces the RCPT, as depicted in Fig. 16g. The decline in partial dependence exhibited an almost linear trend, highlighting a distinct correlation between the increase in the fine aggregate-to-water ratio and the detrimental impact on RCPT. Conversely, changing the fine aggregate ratio to water from 4.2 to 4.9 does not considerably affect the RCPT.

Increasing the silica fume-to-binder ratio within the range of 0–0.06 reduces the RCPT. This is evident from the linear decrease in the partial dependence, as illustrated in Fig. 16h. On the other hand, when the ratio of silica fume to binder is increased from 0.06 to 0.1, a minor increase in the RCPT is observed. As mentioned, the addition of silica fume to the mixtures increases the amounts of C-S-H gel, which fills the pores and improves the microstructure of concrete [81]. However, excessive amounts of pozzolanic materials decelerate cement hydration and create more voids between unreacted SCM particles [78]. The minimum RCPT can be obtained when the silica fume-to-binder ratio is between 0.06 and 0.07.

As shown in Fig. 15i, there is a positive correlation between temperature and RCPT. Increasing the temperature of the mixture increases the RCPT and worsens the durability. Higher temperatures accelerate the hydration process, resulting in a more porous microstructure and lower densities [83], allowing chloride ions to penetrate more easily.

3.4. Managerial and practical implications

There are essential management and practical ramifications to the several facets of this study. From a managerial perspective, the suggested prediction models with high accuracy assist the concrete manufacturers in predicting the durability of concrete. Therefore, projects can be completed faster because powerful machine learning techniques can predict the durability of concrete in a few seconds, which is significantly faster than experimental methods. Further, this study identifies the optimal variables that should be used in the modeling to predict the RCPT of concrete with the highest possible accuracy, which can help modelers apply optimal feature sets in their models.

Additionally, the relative influence of different factors on the RCPT is determined, which can help scholars identify which variables should be considered to significantly improve the RCPT of concrete. Moreover, the presented relation between optimal variables and RCPT can provide concrete specialists and researchers insight into the optimal range of parameters to minimize RCPT. Therefore, they can investigate the values near the presented optimal range of materials to design durable concrete mixtures.

4. Conclusion

This article aims to create models for concrete RCPT prediction and examines the influence of different parameters on RCPT. Feature selection was applied in this research to minimize modeling errors. After that, three modeling methods—XGB, LGBM, and MLP—were used and compared. Efficient hyperparameter tuning remains a significant hurdle in machine learning methodologies. To achieve this, a multi-objective less-parameter meta-heuristic algorithm (STML) was implemented for hyperparameter tuning. The developed model's performance was compared with three GS models. After detecting the best-performing model, the SHAP method was employed to investigate how different parameters affect RCPT, and subsequently, the PDP method was used to identify the optimal value for each parameter.

The results show that STML significantly outperformed GS

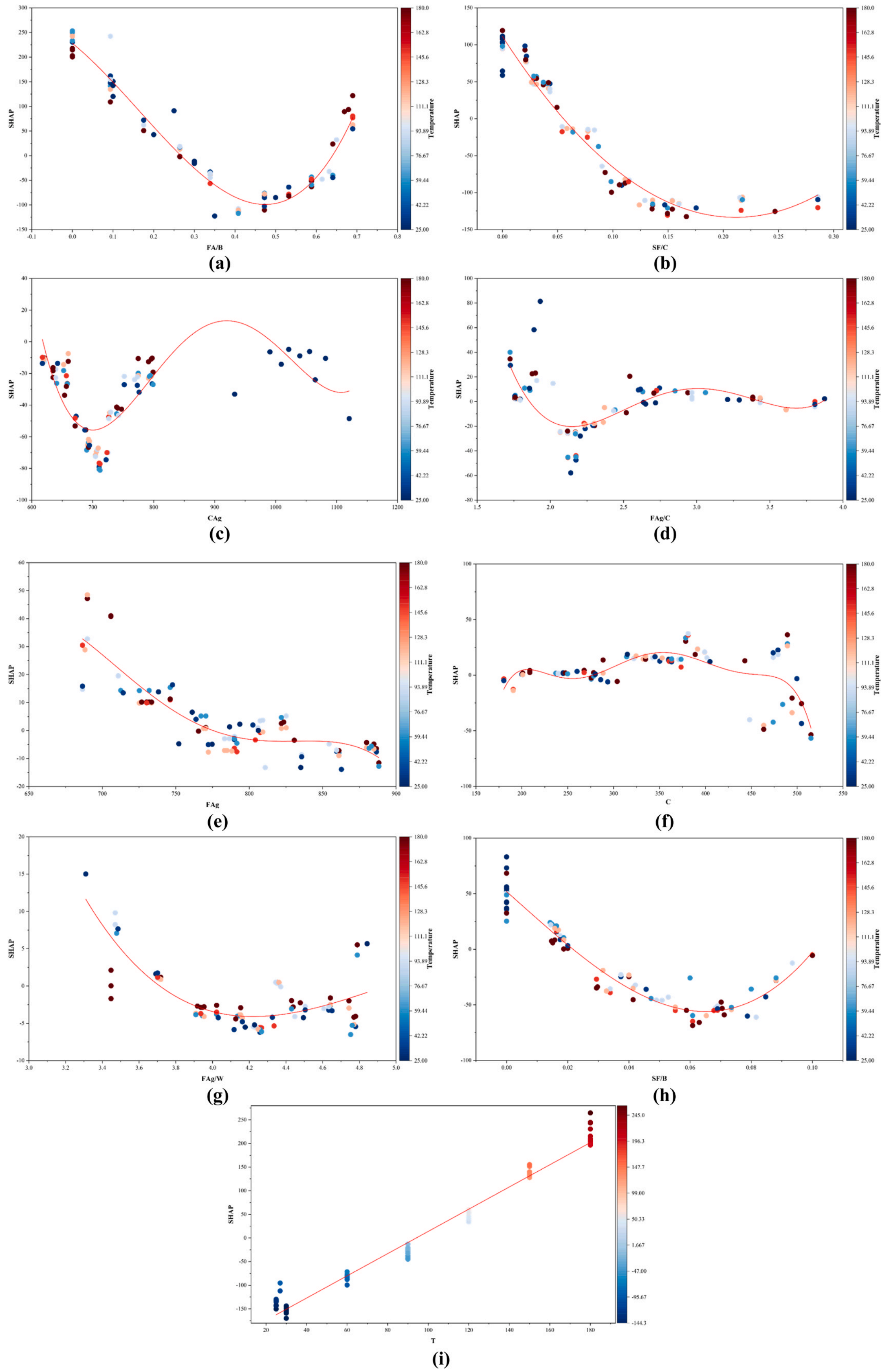

Fig. 16. PDP analysis of the optimal variables for RCPT of concrete.

(conventional hyperparameter tuning method) and XGB and LGBM considerably prevailed in ANN (the conventional method for RCPT prediction). XGB-STML method performed the best for predicting the concrete RCPT compared with other methods. According to the results of XGB-STML, the independent variables in the optimal feature set included the coarse aggregate, fine aggregate to cement ratio, fly ash to binder, temperature, silica fume to cement, cement, fine aggregate, fine aggregate to water, and silica fume to binder. Accordingly, practitioners can apply such variables in developing the RCPT prediction models and predict the RCPT of fly ash concrete in real-life applications. Moreover, the SHAP analysis results revealed that temperature significantly impacts the prediction of RCPT, which was rarely considered in the previous studies, whilst the ratio of fine aggregate to water had the lowest effect.

Achieving optimal durability and performance in concrete requires precise adjustments to material proportions and environmental conditions. This study identified key parameters for minimizing RCPT (Rapid Chloride Permeability Test) and enhancing concrete durability. The optimal fly ash-to-binder ratio was between 0.35 and 0.6, while the silica fume-to-cement ratio should ideally range from 0.15 to 0.3. Furthermore, minimizing RCPT could be achieved by maintaining a coarse aggregate content of approximately 700 kg/m^3 and a fine aggregate-to-cement ratio between 2 and 2.5. Increasing the cement content beyond 450 kg/m^3 also contributed to lower RCPT, as does adjusting the fine aggregate-to-water ratio from 3.3 to around 4.2. Notably, the minimum RCPT was observed when the silica fume-to-binder ratio was maintained between 0.06 and 0.07. However, environmental factors, such as elevated mixture temperatures, adversely affected durability by increasing RCPT. These results underscore the interconnected nature of material composition and environmental conditions, offering valuable insights for developing durable and high-performance concrete.

It is important to clarify that these models can predict the RCPT of observations within the range of the input variables represented in this dataset. Applying the models to observations outside this range and including other materials requires retraining. One of the limitations of this study is that it compares the performance of the developed hyperparameter tuning model with only GS and on a single dataset. Accordingly, it is recommended that future studies compare the performance of the STML-based tuning model with other tuning models on various datasets.

Another limitation of this investigation is applying three performance metrics in tuning the hyperparameters of machine learning techniques. The future studies can apply other metrics in the model and evaluate the influence of adding new metrics on the overall performance of the model. Another limitation of this study is that it applies the Scipy library for modeling the MLP. This library does not let the users perform architectural enhancements (e.g., batch normalization or dropout). Hence, future studies can apply Keras or PyTorch libraries to investigate the influence of architectural enhancements on the performance of MLP. The developed models and the proposed optimal material ranges are valid for any mixture sample within the range of the material shown in Fig. 2. Therefore, they may not be accurate for extrapolation when the range of materials in a sample is not within the range of the applied dataset.

Since this study predicts RCPT values as a measure of chloride ion penetration, it should be mentioned that RCPT has many limitations. RCPT is a rapid test, and it should be validated against long-term ponding tests (ASTM C1543), which require at least 90 days to provide a more accurate and realistic measure of chloride ingress [84,85]. That is, RCPT evaluates the chloride ion penetration based on accelerated ion migration under high voltage (60 V), which can cause heating, moisture loss, and increased ionic mobility. In this regard, this measure (RCPT) can be influenced by factors other than the chloride diffusion [12]. Consequently, when RCPT is predicted from electrical resistivity, these limitations may influence the prediction and interpretation of the results. To address this, we suggest developing prediction models based on the proposed methodology in this study to predict ponding tests in future studies.

Declaration of Generative AI and AI-assisted technologies in the writing process

We have not used generative AI in this study.

CRediT authorship contribution statement

Hamed Naseri: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Farzad Safi Jahanshahi:** Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Amir H. Gandomi:** Writing – review & editing, Supervision, Methodology.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

References

[1] E. Aprianti SA huge number of artificial waste material can be supplementary cementitious material (SCM) for concrete production – a review part II, Elsevier Ltd, 2017..

[2] C. Meyer, The greening of the concrete industry, Cem. Concr. Compos. 31 (2009) 601–605, https://doi.org/10.1016/J.CEMCONCOMP.2008.12.010.

[3] P.K. Mehta, Greening of the concrete industry for sustainable development, Concr. Int. 24 (2002) 23–28.

[4] R. Feiz, J. Ammenberg, M. Eklund, A. Helgstrand, R. Marshall, Improving the CO2 performance of cement, part I: utilizing life-cycle assessment and key performance indicators to assess development within the cement industry, J. Clean. Prod. 98 (2015) 272–281, https://doi.org/10.1016/J.JCLEPRO.2014.01.083.

[5] H. Naseri, H. Jahanbakhsh, F.Moghadas Nejad, Developing a novel machine learning method to predict the compressive strength of Fly ash concrete in different ages, AUT J. Civ. Eng. 0 (2019), https://doi.org/10.22060/AJCE.2019.16124.5569.

[6] P. Van Den Heede, N. De Belie, Environmental impact and life cycle assessment (LCA) of traditional and "Green" concretes: literature review and theoretical calculations, Cem. Concr. Compos. 34 (2012) 431–442, https://doi.org/10.1016/j.cemconcomp.2012.01.004.

[7] C.L. Page, M.M. PageDurability of concrete and cement composites, Elsevier, 2007. https://doi.org/10.1533/9781845693398..

[8] L. Cheng, H. Jin, J. Liu, J. Liu, J. Zhu, F. Xing, A novel approach for determining the chloride ion diffusion coefficient in seawater sand concrete: the extravasation method, Constr. Build. Mater. 441 (2024) 137454, https://doi.org/10.1016/j.conbuildmat.2024.137454.

[9] H.W. Song, S.J. Kwon, Evaluation of chloride penetration in high performance concrete using neural network algorithm and micro pore structure, Cem. Concr. Res. 39 (2009) 814–824, https://doi.org/10.1016/j.cemconres.2009.05.013.

[10] C. Andrade, R. D'Andrea, N. Rebolledo, Chloride ion penetration in concrete: the reaction factor in the electrical resistivity model, Cem. Concr. Compos. 47 (2014) 41–46, https://doi.org/10.1016/j.cemconcomp.2013.09.022.

[11] J. Shi, B. Liu, J. Qin, J. Jiang, X. Wu, J. Tan, Experimental study of performance of repair mortar: evaluation of in-situ tests and correlation analysis, J. Build. Eng. 31 (2020) 101325, https://doi.org/10.1016/j.jobe.2020.101325.

[12] D. Whiting, Rapid measurement of the chloride permeability of concrete, Public Roads 45 (1981) 101–112.

[13] M.D.A. Thomas, P.B. Bamforth, Modelling chloride diffusion in concrete effect of Fly ash and slag, Cem. Concr. Res. 29 (1999) 487–495, https://doi.org/10.1016/S0008-8846(98)00192-6.

[14] S. Kumar, B. Rai, R. Biswas, P. Samui, D. Kim, Prediction of rapid chloride permeability of self-compacting concrete using multivariate adaptive regression spline and minimax probability machine regression, J. Build. Eng. 32 (2020) 101490, https://doi.org/10.1016/j.jobe.2020.101490.

[15] O. Habibi, T. Youssef, H. Naseri, K. Ibrahim, Ensemble learning models for prediction of punching shear strength in RC slab-column connections, Civ. Eng. J. 10 (2024) 1–20, https://doi.org/10.28991/CEJ-SP2024-010-01.

[16] J. Morgan, R. Daugherty, A. Hilchie, B. Carey, Sample size and modeling accuracy with decision tree based data mining tools, Acad. Inf. Manag. Sci. J. 6 (2003) 71–99.

[17] M.Z. Joharestani, C. Cao, X. Ni, B. Bashir, S. Talebiesfandarani, PM2.5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data, Atmosphere 10 (2019) 373, https://doi.org/10.3390/atmos10070373.

[18] W. Cai, R. Wei, L. Xu, X. Ding, A method for modelling greenhouse temperature using gradient boost decision tree, Inf. Process. Agric. 9 (2022) 343–354, https://doi.org/10.1016/j.inpa.2021.08.004.

[19] A.R. Boğa, M. Öztürk, I.B. Topçu, Using ANN and ANFIS to predict the mechanical and chloride permeability properties of concrete containing GGBFS and CNI, Compos. Part B Eng. 45 (2013) 688–696, https://doi.org/10.1016/J.COMPOSITESB.2012.05.054.

[20] M. Najimi, N. Ghafoori, M. Nikoo, Modeling chloride penetration in self-consolidating concrete using artificial neural network combined with artificial bee colony algorithm, J. Build. Eng. 22 (2019) 216–226, https://doi.org/10.1016/j.jobe.2018.12.013.

[21] N. Ghafoori, M. Najimi, J. Sobhani, M.A. Aqel, Predicting rapid chloride permeability of self-consolidating concrete: a comparative study on statistical and neural network models, Constr. Build. Mater. 44 (2013) 381–390, https://doi.org/10.1016/j.conbuildmat.2013.03.039.

[22] O.A. Mohamed, M. Ati, W. Al Hawat, Implementation of artificial neural networks for prediction of chloride penetration in concrete, Int. J. Eng. Technol. 7 (2018) 47–52, https://doi.org/10.14419/ijet.v7i2.28.12880.

[23] O. Mohamed, M. Kewalramani, M. Ati, W.Al Hawat, Application of ANN for prediction of chloride penetration resistance and concrete compressive strength, Materialia 17 (2021) 101123, https://doi.org/10.1016/j.mtla.2021.101123.

[24] V.Q. Tran, V.L. Giap, D.P. Vu, R.C. George, L.S. Ho, Application of machine learning technique for predicting and evaluating chloride ingress in concrete, Front. Struct. Civ. Eng. 16 (2022) 1153–1169, https://doi.org/10.1007/s11709-022-0830-4.

[25] D.M. Ge, L.C. Zhao, M. Esmaeili-Falak, Estimation of rapid chloride permeability of SCC using hyperparameters optimized random forest models, J. Sustain. Cem. Mater. 12 (2023) 542–560, https://doi.org/10.1080/21650373.2022.2093291.

[26] A.-T. Tran, T.-H. Le, H.M. Nguyen, Forecast of surface chloride concentration of concrete utilizing ensemble decision tree boosted, J. Sci. Transp. Technol. 2 (2022) 44–56, https://doi.org/10.58845/jstt.utt.2022.en57.

[27] R. Kazemi, A. Gholampour, Evaluating the rapid chloride permeability of self-compacting concrete containing Fly ash and silica fume exposed to different temperatures: an artificial intelligence framework, Constr. Build. Mater. 409 (2023) 133835, https://doi.org/10.1016/j.conbuildmat.2023.133835.

[28] A. Abdulalim Alabdullah, M. Iqbal, M. Zahid, K. Khan, M. Nasir Amin, F.E. Jalal, Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis, Constr. Build. Mater. 345 (2022) 128296, https://doi.org/10.1016/j.conbuildmat.2022.128296.

[29] M.N. Amin, M. Raheel, M. Iqbal, K. Khan, M.G. Qadir, F.E. Jalal, A.A. Alabdullah, A. Ajwad, M.A. Al-Faiad, A.M. Abu-Arab, Prediction of rapid chloride penetration resistance to assess the influence of affecting variables on Metakaolin-Based concrete using gene expression programming, Materials 15 (2022) 6959, https://doi.org/10.3390/ma15196959.

[30] E. Güneyisi, M. Gesoğlu, T. Özturan, E. Özbay, Estimation of chloride permeability of concretes by empirical modeling: considering effects of cement type, curing condition and age, Constr. Build. Mater. 23 (2009) 469–481, https://doi.org/10.1016/j.conbuildmat.2007.10.022.

[31] O.A. Hodhod, H.I. Ahmed, Developing an artificial neural network model to evaluate chloride diffusivity in high performance concrete, HBRC J. 9 (2013) 15–21, https://doi.org/10.1016/j.hbrcj.2013.04.001.

[32] M.I. Khan, Predicting properties of high performance concrete containing composite cementitious materials using artificial neural networks, Autom. Constr. (2012) 516–524, https://doi.org/10.1016/j.autcon.2011.11.011.

[33] R. Parichatprecha, P. Nimityongskul, Analysis of durability of high performance concrete using artificial neural networks, Constr. Build. Mater. 23 (2009) 910–917, https://doi.org/10.1016/j.conbuildmat.2008.04.015.

[34] P. Liashchynskyi, P. LiashchynskyiGrid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS, ArXiv Prepr. 2019. https://doi.org/doi.org/10.48550/arXiv.1912.06059..

[35] T. Hai, A. Basem, A. Alizadeh, K. Sharma, D.J. Jasim, H. Rajab, M. Ahmed, M. Kassim, N.S.S. Singh, H. Maleki, Optimizing Gaussian process regression (GPR) hyperparameters with three metaheuristic algorithms for viscosity prediction of suspensions containing microencapsulated PCMs, Sci. Rep. 14 (2024) 1–24, https://doi.org/10.1038/s41598-024-71027-9.

[36] H. Naseri, E.O.D. Waygood, B. Wang, Z. Patterson, Application of machine learning to child mode choice with a novel technique to optimize hyperparameters, Int. J. Environ. Res. Public Health 19 (2022) 16844, https://doi.org/10.3390/ijerph192416844.

[37] B. Zolghadr-Asli, M. Latifi, R. Beig Zali, M.R. Nikoo, R. Farmani, R. Nazari, A. H. Gandomi, Self-tuning multi-layer optimization algorithm (STML): an innovative parameter-less approach, Appl. Soft Comput. 165 (2024) 112045, https://doi.org/10.1016/j.asoc.2024.112045.

[38] H. Naseri, E.O.D. Waygood, Z. Patterson, Household transportation lifecycle greenhouse gas emission prediction, Transp. Res. Part D. Transp. Environ. 141 (2025) 104660, https://doi.org/10.1016/j.trd.2025.104660.

[39] S.S. More, A.R. Kambekar, Experimental investigation of durable parameters in high volume flyash concrete used in coastal region. IOP Conf. Ser. Earth Environ. Sci., IOP Publishing, 2024 012044, https://doi.org/10.1088/1755-1315/1326/1/012044.

[40] S. Kumar, D.R. Kumar, W. Wipulanusat, S. Keawsawasvong, Development of ANN-based metaheuristic models for the study of the durability characteristics of high-volume Fly ash self-compacting concrete with silica fume, J. Build. Eng. 94 (2024) 109844, https://doi.org/10.1016/j.jobe.2024.109844.

[41] M.K. Rao, D. Naga, S. Kumar, Durability assessment of concrete with Class-F Fly ash by chloride ion permeability, Int. J. Recent Technol. Eng. 8 (2019) 8831–8836, https://doi.org/10.35940/ijrte.d9470.118419.

[42] B.B. Das, S.P. Pandey, Influence of fineness of fly ash on the carbonation and electrical conductivity of concrete, J. Mater. Civ. Eng. (2011), https://doi.org/10.1061/(ASCE)MT.1943-5533.0000298.

[43] H. Naseri, Cost optimization of no-slump concrete using genetic algorithm and particle swarm optimization, Int. J. Innov. Manag Technol. 10 (2019), https://doi.org/10.18178/ijimt.2019.10.1.832.

[44] H. Naseri, P. Hosseini, H. Jahanbakhsh, P. Hosseini, A.H. Gandomi, A novel evolutionary learning to prepare sustainable concrete mixtures with supplementary cementitious materials, Environ. Dev. Sustain. 2022 (2022) 1–35, https://doi.org/10.1007/S10668-022-02283-W.

[45] H. Naseri, H. Jahanbakhsh, K. Khezri, A.A. Shirzadi JavidToward sustainability in optimizing the fly ash concrete mixture ingredients by introducing a new prediction algorithm, Springer Netherlands, 2021. https://doi.org/10.1007/s10668-021-01554-2..

[46] H. Naderpour, A.H. Rafiean, P. Fakharian, Compressive strength prediction of environmentally friendly concrete using artificial neural networks, J. Build. Eng. 16 (2018) 213–219, https://doi.org/10.1016/j.jobe.2018.01.007.

[47] H. Naseri, E.O.D. Waygood, B. Wang, Z. Patterson, R.A. Daziano, A novel feature selection technique to better predict climate change stage of change, Sustainability 14 (2021) 40, https://doi.org/10.3390/su14010040.

[48] H. Naseri, M. Shokoohi, H. Jahanbakhsh, M.M. Karimi, E.O.D. Waygood, Novel Soft-Computing approach to better predict flexible pavement roughness, Transp. Res. Rec. J. Transp. Res. Board (2023) 036119812311610, https://doi.org/10.1177/03611981231161051.

[49] M. Awad, S. Fraihat, Recursive feature elimination with cross-validation with decision tree: feature selection method for machine learning-based intrusion detection systems, J. Sens. Actuator Netw. 12 (2023) 67, https://doi.org/10.3390/jsan12050067.

[50] H. Naseri, E.O.D. Waygood, Z. Patterson, B. Wang, Which variables influence electric vehicle adoption? Transportation (2024) 1–38, https://doi.org/10.1007/s11116-024-10525-1.

[51] H. Naseri, E.O.D. Waygood, Z. Patterson, B. Wang, Who is more likely to buy electric vehicles? Transp. Policy 155 (2024) 15–28, https://doi.org/10.1016/J.TRANPOL.2024.06.013.

[52] A. Sharma, B. Singh, AE-LGBM: Sequence-based novel approach to detect interacting protein pairs via ensemble of autoencoder and LightGBM, Comput. Biol. Med. 125 (2020) 103964, https://doi.org/10.1016/j.compbiomed.2020.103964.

[53] Y. Guo, H. Ma, S. Tian, Quantitative investment model based on LightGBM algorithm, Acad. J. Comput. Inf. Sci. 5 (2022) 40–46, https://doi.org/10.25236/ajcis.2022.050707.

[54] Y. Liu, C. Lyu, A. Khadka, W. Zhang, Z. Liu, Spatio-temporal ensemble method for car-hailing demand prediction, IEEE Trans. Intell. Transp. Syst. 21 (2020) 5328–5333, https://doi.org/10.1109/TITS.2019.2948790.

[55] L. Tutica, K. Vineel, S. Mishra, M.K. Mishra, S. Suman, Invoice deduction classification using LGBM prediction model. Lect. Notes Electr. Eng., Springer Science and Business Media Deutschland GmbH, 2021, pp. 127–137, https://doi.org/10.1007/978-981-15-8752-8_13.

[56] Q. Shangguan, T. Fu, J. Wang, S. Fang, L. Fu, A proactive lane-changing risk prediction framework considering driving intention recognition and different lane-changing patterns, Accid. Anal. Prev. 164 (2022) 106500, https://doi.org/10.1016/j.aap.2021.106500.

[57] X. Guo, Y. Gao, D. Zheng, Y. Ning, Q. Zhao, Study on short-term photovoltaic power prediction model based on the stacking ensemble learning, Energy Rep. 6 (2020) 1424–1431, https://doi.org/10.1016/j.egyr.2020.11.006.

[58] H. Naseri, E.O.D. Waygood, B. Wang, Z. Patterson, Interpretable machine learning approach to predicting electric vehicle buying decisions, Transp. Res. Rec. J. Transp. Res. Board 036119812311695 (2023), https://doi.org/10.1177/03611981231169533.

[59] S. Ghavami, H. Naseri, F.S. Jahanshahi, Enhanced prediction and uncertainty modeling of pavement roughness using machine learning and conformal prediction, Infrastructures 10 (2025) 166, https://doi.org/10.3390/INFRASTRUCTURES10070166.

[60] H. Naseri, F. Ciari, C.A. Isler, Influence of pedestrianization on travel behavior changes: a case study of Montreal, J. Transp. Health 44 (2025) 102123, https://doi.org/10.1016/j.jth.2025.102123.

[61] H. Naseri, A. Aliakbari, M.A. Javadian, A. Aliakbari, E.O.D. Waygood, A novel technique for multi-objective sustainable decisions for pavement maintenance and rehabilitation, Case Stud. Constr. Mater. 20 (2024) e03037, https://doi.org/10.1016/J.CSCM.2024.E03037.

[62] T. Nguyen-Sy, J. Wakim, Q.D. To, M.N. Vu, T.D. Nguyen, T.T. Nguyen, Predicting the compressive strength of concrete from its compositions and age using the extreme gradient boosting method, Constr. Build. Mater. 260 (2020) 119757, https://doi.org/10.1016/j.conbuildmat.2020.119757.

[63] P. Salas, R. De la Fuente, S. Astroza, J.A. Carrasco, A systematic comparative evaluation of machine learning classifiers and discrete choice models for travel

mode choice in the presence of response heterogeneity, Expert Syst. Appl. 193 (2022) 116253, https://doi.org/10.1016/j.eswa.2021.116253.
[64] M.K. Hasan, M.A. Alam, D. Das, E. Hossain, M. Hasan, Diabetes prediction using ensembling of different machine learning classifiers, IEEE Access (2020), https://doi.org/10.1109/ACCESS.2020.2989857.
[65] R.V. Rao, V.J. Savsani, D.P. Vakharia, Teaching-learning-based optimization: a novel method for constrained mechanical design optimization problems, CAD Comput. Aided Des. 43 (2011) 303–315, https://doi.org/10.1016/j.cad.2010.12.015.
[66] A.R. Ghanizadeh, F. Safi Jahanshahi, S.S. Naseralavi, Intelligent modelling of unconfined compressive strength of cement stabilised iron ore tailings: a case study of golgohar mine, Eur. J. Environ. Civ. Eng. 28 (2024) 1759–1787, https://doi.org/10.1080/19648189.2023.2276133.
[67] P. Fakharian, Y. Nouri, A.R. Ghanizadeh, F. Safi Jahanshahi, H. Naderpour, A. Kheyroddin, Bond strength prediction of externally bonded reinforcement on groove method (EBROG) using Mars-POA, 118532, Compos. Struct. (2024) 349–350, https://doi.org/10.1016/j.compstruct.2024.118532.
[68] F. Safi Jahanshahi, A.R. Ghanizadeh, Machine learning approaches for resilient modulus modeling of cement-stabilized magnetite and hematite iron ore tailings, Sci. Rep. 15 (2025) 1–37, https://doi.org/10.1038/s41598-025-86978-w.
[69] Y. Nouri, A.R. Ghanizadeh, F. Safi Jahanshahi, P. Fakharian, Data-driven prediction of axial compression capacity of GFRP-reinforced concrete column using soft computing methods, J. Build. Eng. 101 (2025) 111831, https://doi.org/10.1016/j.jobe.2025.111831.
[70] S.M. Lundberg, S.I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. (2017) 4766–4775. ⟨https://github.com/slundberg/shap⟩ (accessed June 18, 2022).
[71] C. Cakiroglu, S. Demir, M. Hakan Ozdemir, B. Latif Aylak, G. Sariisik, L. Abualigah, Data-driven interpretable ensemble learning methods for the prediction of wind turbine power incorporating SHAP analysis, Expert Syst. Appl. 237 (2024) 121464, https://doi.org/10.1016/j.eswa.2023.121464.
[72] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. 29 (2001) 1189–1232, https://doi.org/10.1214/aos/1013203451.
[73] V. Quan Tran, V. Quoc Dang, L. Si Ho, Evaluating compressive strength of concrete made with recycled concrete aggregates using machine learning approach, Constr. Build. Mater. 323 (2022) 126578, https://doi.org/10.1016/j.conbuildmat.2022.126578.
[74] A. Goldstein, A. Kapelner, J. Bleich, E. Pitkin, Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation, J. Comput. Graph. Stat. 24 (2015) 44–65, https://doi.org/10.1080/10618600.2014.907095.
[75] B. Sankar, P. Ramadoss, Modelling the compressive strength of high-performance concrete containing metakaolin using distinctive statistical techniques, Results Control Optim. 12 (2023) 100241, https://doi.org/10.1016/j.rico.2023.100241.
[76] B.S. Dhanya, M. Santhanam, Performance evaluation of rapid chloride permeability test in concretes with supplementary cementitious materials, Mater. Struct. Constr. 50 (2017) 1–9, https://doi.org/10.1617/s11527-016-0940-3.
[77] S. Arunchaitanya, S. Dey, Sorptivity and rapid chloride ion penetration of self-compacting concrete using Fly ash and copper slag, AI Civ. Eng. 2 (2023) 1–14, https://doi.org/10.1007/s43503-023-00013-3.
[78] L.S. Ho, T.P. Huynh, Long-term mechanical properties and durability of high-strength concrete containing high-volume local Fly ash as a partial cement substitution, Results Eng. 18 (2023) 101113, https://doi.org/10.1016/j.rineng.2023.101113.
[79] A.R. Bagheri, H. Zanganeh, M.M. Moalemi, Mechanical and durability properties of ternary concretes containing silica fume and low reactivity blast furnace slag, Cem. Concr. Compos. 34 (2012) 663–670, https://doi.org/10.1016/j.cemconcomp.2012.01.007.
[80] S.K. Chaudhary, A.K. Sinha, Effect of silica fume on permeability and microstructure of high strength concrete, Civ. Eng. J. 6 (2020) 1697–1703, https://doi.org/10.28991/cej-2020-03091575.
[81] A. Cuesta, I. Santacruz, A.G. De la Torre, M. Dapiaggi, J.D. Zea-Garcia, M.A. G. Aranda, Local structure and Ca/Si ratio in C-S-H gels from hydration of blends of tricalcium silicate and silica fume, Cem. Concr. Res. 143 (2021) 106405, https://doi.org/10.1016/j.cemconres.2021.106405.
[82] X. Liu, T. Li, W. Tian, Y. Wang, Y. Chen, Study on the durability of concrete with FNS fine aggregate, J. Hazard. Mater. 381 (2020) 120936, https://doi.org/10.1016/j.jhazmat.2019.120936.
[83] F. Vodák, K. Trtík, O. Kapičková, Š. Hošková, P. Demo, The effect of temperature on strength - porosity relationship for concrete, Constr. Build. Mater. 18 (2004) 529–534, https://doi.org/10.1016/j.conbuildmat.2004.04.009.
[84] ASTM C1202, Standard Test Method for Electrical Indication of Concrete's Ability to Resist Chloride Ion Penetration, Am. Soc. Test. Mater. 2012. pp.1–8. https://doi.org/10.1520/C1202-22.
[85] ASTM C1543-10a, C1543 standard test method for determining the penetration of chloride ion into concrete by ponding (Withdrawn 2019), ASTM 2019. ⟨https://store.astm.org/c1543-10a.html⟩ (accessed July 18, 2025).