



# Prediction of surface chloride concentration of marine concrete using ensemble machine learning

Rong Cai<sup>a,b</sup>, Taihao Han<sup>c,1</sup>, Wenyu Liao<sup>a</sup>, Jie Huang<sup>d</sup>, Dawang Li<sup>e</sup>, Aditya Kumar<sup>c</sup>, Hongyan Ma<sup>a,\*</sup>

<sup>a</sup> Department of Civil, Architectural and Environmental engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA

<sup>b</sup> Department of Construction, Guangxi University of Finance and Economics, Nanning 530008, China

<sup>c</sup> Department of Materials Science and Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA

<sup>d</sup> Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA

<sup>e</sup> Guangdong Province Key Laboratory of Durability for Marine Civil Engineering, School of Civil Engineering, Shenzhen University, Shenzhen, Guangdong 518060, China

## ARTICLE INFO

### Keywords:

Surface chloride concentration  
Marine concrete  
Ensemble machine learning  
Voting method

## ABSTRACT

This paper develops and employs an ensemble machine learning (ML) model for prediction of surface chloride concentration ( $C_s$ ) of concrete, which is an essential parameter for durability design and service life prediction of concrete structures in marine environment. For this purpose, a database containing 642 data-records of field exposure data of  $C_s$  (along with the associated mixture proportion parameters, environmental conditions and exposure time) is established based on extensive literature surveying, which covers splash, tidal, and submerged zones in various areas in the world. The database is used to train five standalone ML models, that is, linear regression (LR), Gaussian process regression (GPR), support vector machine (SVM), multilayer perceptron artificial neural network (MLP-ANN) and random forests (RF) models, as well as an ensemble weighted voting-based ML model, and subsequently used to compare their prediction performances. It is shown that, by meta-heuristically combining predictions of RF, MLP-ANN, and SVM, the ensemble ML model produces higher accuracy of prediction compared to all standalone ML models tested in this study. The prediction performances of eight conventional quantitative models for  $C_s$  prediction are also analyzed based on the testing dataset selected for ML. The results show that adoption of more diverse datasets and consideration of more factors in conventional models can improve their prediction performance. The ensemble ML model established on a large database, can easily consider the twelve influencing factors (which is difficult for conventional models) in the database, and has superior prediction performance, yet better time-efficiency, compared to conventional models.

## 1. Introduction

Reinforced concrete (RC) structures are widely used in engineering practice in marine and coastal environments, such as construction of sea-crossing bridges, harbor docks, coastal roads and buildings. Reinforcing steel in RC structures can generally be protected against corrosion by a passive oxide film, which is quite stable in the highly alkaline micro-environment provided by the pore solution of concrete [1–3]. However, chloride ions from seawater and atmosphere (in the form of aerosol) in marine/coastal environment can deposit on the surface of RC structures and penetrate into concrete. If chloride ions reach and accumulate in the concrete matrix surrounding the reinforcing steel, they could destruct the passive film, initiate and accelerate steel corrosion, leading to cracking and spalling of concrete and degradation of the load-carrying capacity of RC structures [4,5]. This

process implies progressively significant deterioration with increased exposure time [6]. As a result, some RC structures fail to complete the predetermined or designed service life and damaged prematurely. Corrosion of steel in RC structures not only affects normal functions of engineering structures, but also results in engineering accidents, safety hazard, and great economic losses [1,7–9]. Therefore, durability design or service life prediction has become one of the major tasks in the present design of concrete structures. This task must be based on consistent models that can account for various influential factors and describe the deterioration mechanisms more accurately [10].

According to the action classes and characteristics of chloride exposure, marine environments are usually classified into four zones, that is, atmospheric, tidal, splash, and submerged zones [11,12]. In the submerged zone, concrete can be assumed to be saturated, and the transport of chloride ions in concrete cover is driven by pure diffusion.

\* Corresponding author.

E-mail address: [mahon@mst.edu](mailto:mahon@mst.edu) (H. Ma).

<sup>1</sup> Contributed equally to the first author.

In other zones, such as the tidal zone and the splash zone, absorption is an additional transport mechanism that facilitates chloride ingress. In the atmospheric zone, the transport mechanism and path of chloride ions are more complex than other zones [6], because of factors such as distance to sea front and direction/speed of wind. Therefore, this paper does not include the atmospheric zone, but rather focuses on the other three zones.

In general, no matter in which zone, Fick's second law of diffusion is used to describe the overall ingress of chloride into concrete [13,14]. The analytical solution of Fick's second law is given in Eq. (1), which has been widely adopted in service life design of RC structures in marine environments [15].

$$C(x, t) = C_0 + (C_s - C_0) \left[ 1 - \operatorname{erf} \left( \frac{x}{2\sqrt{D \cdot t}} \right) \right] \quad (1)$$

where  $C(x, t)$  is chloride concentration in distance  $x$  from the surface after exposure time  $t$ ;  $C_0$  is the initial chloride concentration in concrete;  $x$  is the depth from the exposed concrete surface;  $D$  is the apparent chloride diffusion coefficient;  $C_s$  is the apparent surface chloride concentration;  $\operatorname{erf}(\cdot)$  is the error function. For any given concrete,  $C_0$  is a constant, and chloride ingress is thus determined by  $C_s$  and  $D$ . The chloride diffusion coefficient  $D$  is a time-dependent material property, and it can be determined/predicted on the bases of information pertaining to the microstructure and material composition. As compared to  $D$ ,  $C_s$  is a more complicated variable, since it depends on not only material properties, but also time and environmental features. Therefore, more studies are needed to understand the buildup of  $C_s$ , so as to predict this parameter and chloride ingress more accurately.

This paper focuses on improving predictions of  $C_s$  values. Database used in this work consists of 642 data-records of the apparent surface chloride concentration in marine concrete, collected from past publications. All data-records included in the database are from field measurements on RC structures located in marine splash, tidal, and submerged zones. Specifically, they are obtained from curve fitting – based on Eq. (1) – of chloride penetration profiles, as illustrated in Fig. 1. In most field concretes, a convection zone is present, as shown in Fig. 1(a). In this case, the bulk diffusion section of the chloride profile is fitted to obtain  $C_s$ . Some lab-based tests may not show convection zone, so the whole profile can be fitted for determination of  $C_s$ , as illustrated in Fig. 1(b).

According to Eq. (1), the apparent surface chloride concentration  $C_s$  is an essential parameter, since it not only represents the intensity of aggressive action of the service environment, but also provides the boundary condition for service life prediction and quantitative durability design of RC structures in marine/coastal environment [17,18].

Although guidelines tend to assume  $C_s$  a constant in a given environment when Eq. (1) is employed, this assumption may lead to significant error. Many past publications have reported highly diverse field values of  $C_s$  and provided quantitative models for this important parameter [6,12,16,18–25]. However, the models have not shown good performance in predicting or describing  $C_s$ , given that  $C_s$  is a complex parameter determined by so many factors, such as environmental factors (chloride concentration of sea water, zonation, action of carbonation, temperature, relative humidity, etc.), material factors (binder content, binder composition, water-to-binder ratio, etc.), and exposure time [16,19,20,24,26–37]. References [10,29,38–41] have proposed time-variant models on  $C_s$  by using logarithmic, power, exponential and other functions, respectively, but they cannot describe reasonably the development rules according to which  $C_s$  values increase rapidly in the early stage and tend to become stable in the later stage. Moreover, those models neglected other important factors, that is, material composition and environment action classes. References [16, 34, 42] have established models to predict  $C_s$  in relation to water-to-binder ratio ( $w/b$ ) and exposure time  $t$ , but neglected impact of binder type and environmental factors. References [12, 20] provided models of  $C_s$  with variables of materials and environment action classes, but neglected the crucial impact of exposure time. It can be seen that these conventional quantitative  $C_s$  models can only consider some – not all – of the influential factors, because of the lack of a large amount of experimental data and powerful approach to consider so many variables.

To resolve the complexity of  $C_s$  prediction, the present paper seeks to develop a new method, which can be used to predict/select this key parameter effectively, thus aiding in durability design of RC structures in marine environments. Machine learning (ML), a branch of artificial intelligence, has been successfully used in civil engineering [43,44] and for prediction of corrosion [45–47] by using nonlinear independent variables as input. To compensate for the drawbacks of conventional methods – used for prediction of chloride ingress in concrete – supervised ML models are applied to develop accurate and effective models. The usage of ML models overcomes the complicated attributes of concrete which are exceedingly large compositional degrees of freedom and the inherent nonlinear relationship between mixture independent variable and properties. Recently, some published papers have focused on corrosion of concrete via ML models [48–50], but most of them only consider single ML or statistical techniques, homogenous concrete composition, or simplex environment. It is unknown if ensemble ML models – designed by unifying two or more separate ML models into one – can perform better than single models in different corrosion environments.

In this paper, six ML models – linear regression, Gaussian process regression, multilayer perceptron artificial neural network, random

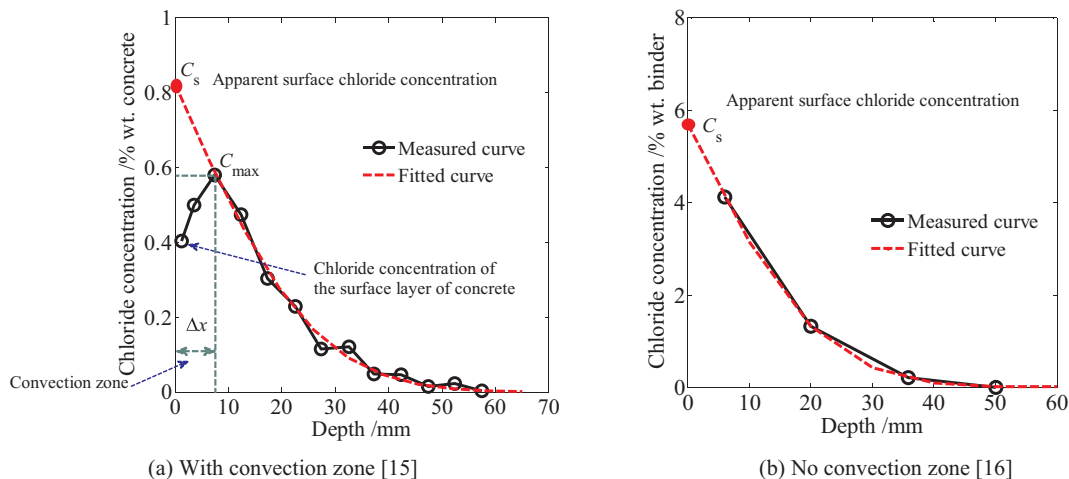


Fig. 1. Schematic diagrams of determination of  $C_s$  in different conditions.

forest, support vector machine, and weighted voting-based ensemble ML – are used for prediction of  $C_s$  in concrete upon exposure to tidal, splash, and submerged zones. Prediction performances of conventional methods, standalone ML models, and the ensemble ML model (which combines multiple ML models) are compared to identify the best model for predicting the  $C_s$  of concrete. Prediction performance of the ML models are rigorously examined in terms of five different statistical parameters – correlation coefficient ( $r$ ), coefficient of determination ( $R^2$ ), mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) – as well as a composite performance index. The comparisons with conventional quantitative  $C_s$  models based on field experimental data have validated the advantages and applicability of the proposed ML method.

## 2. Machine learning models

This section presents a concise overview of six ML models – five standalone models and one ensemble model – that have been implemented in this study. Further details pertaining to the ML models are provided in the Supplementary information 1.

Linear regression (LR) uses piecewise linear functions, which are driven by independent predictors, to predict a numerical target based on a set of inputs [51]. In order to minimize the deviation between the predicted output and the measured value, optimum values of the weight constant and the regression coefficients are determined based on the minimization of mean square error (MSE). In this study, the Nelder-Mead based simplex method [52] was used for minimization of the MSE. Gaussian process regression (GPR) is a regression method that employs a stochastic process to collect random variables, any finite number of which has a joint Gaussian distribution [53]. GPR organizes data in a manner that any given subset of the organized data invariably follows a multivariate Gaussian distribution. For a realistic dataset (e.g., those pertaining to heterogeneous materials such as concrete), a Gaussian noise term that reflects the randomness in the observations is incorporated in the GPR model [54]. The accuracy of the predictions of GPR model depends heavily on the kernel function, which is used to express the covariance. The Pearson VII function was chosen as the kernel function in this study. Artificial neural network (ANN) consists of a layer-based arrangement of multiple computational elements (termed as neurons) resembling the network of neurons in the human brain that hierarchically process and propagate information [55]. Multilayer perceptron artificial neural network (MLP-ANN) is a subclass of ANN with strong self-learning capabilities [56]. In the MLP-ANN model used in this study, the neural network architecture comprised of 5 hidden layers, wherein each layer comprised of  $(2m + 1)$  neurons [57];  $m$  is the number of input variables of the training dataset. Support vector machine (SVM) is a ML methodology for approximating the nonlinear relationship between input variables and output of a dataset by using an optimization approach – rather than a regression approach – to minimize a cost function, or simply put, to transform input data into a higher-dimensional structure such that data with similar characteristics are sequestered from dissimilar ones [58,59].

Random Forest (RF), a collection of tree predictors, is based on two machine learning techniques: bagging and random feature selection [60,61]. The RF model deploys a large number of independent trees that encompass a subset of the training data in a homogenized manner [62] and splits each tree node with a subset feature. The RF model can be summarized in the following steps:

- “ $n_t$ ” bootstrap samples are generated randomly from the training set, and the number of bootstrap samples is equal to the number of trees.
- Grow unpruned regression trees for each of the  $n_t$  bootstrap datasets. The number of leaves of each tree is held constant across the entire model.
- Next, each of the  $n_t$  trees is utilized to predict a data-point outside of the selected bootstrap space. The output of the prediction is

designated as out of bag (OOB) prediction [62].

- All OOB predictions are subsequently aggregated and averaged to produce the overall OOB prediction and OOB error rate.

Voting is an ensemble method of combining predictions from two or more ML models into one. The simplest form of voting – commonly referred to as majority voting – involves invoking each of the ML models to make individual predictions based on data-records from the training set [63–65]. Each prediction made by the ML model counts as a vote, which serves as a measure of its prediction accuracy. The unified prediction is decided by the majority of the votes, entailing that the ML model with the most votes is selected for the final prediction. In this paper, the weighted voting method with the combination of RF, MLP, and SVM models was used as the ensemble ML model.

The authors would like to point out that all optimal hyper-parameters of each model and each kernel function were determined by 10-fold cross-validation (CV) method [66–68]. In short, the 10-fold CV method randomly splits the training database into 10 folds equally. The ML model and the hyper-parameters/functions of which that need to be optimized are 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 the training of the ML model and its blind-testing. With each training-followed-by-testing iteration, the CV error is estimated, and on such basis, the relevant parameters of the ML models are progressively fine-tuned.

## 3. Data collection and performance evaluation of machine learning models

### 3.1. Data collection

The  $C_s$  dataset used in this study consists of a total of 642 sets of data, all of which are field data consolidated from published references. The data includes 386 records in tidal zone [16,19,21–23,30,69–78], 122 records in splash zone [19,21,24,36,69–71,77,78], 134 records in submerged zone [22,24,26,69,70,77,79,80]. The full dataset has been included in the Supplementary information 2. A large portion of the data were used to train the ML models (described in Section 2 and the Supplementary information 1), and the rests (untrained data-domains) were employed to assess their prediction performance.

The  $C_s$  dataset pertains to surface chloride concentration ( $C_s$ , % by weight of concrete) in relation to variations in concrete mixture designs, environmental conditions and exposure time. The values of  $C_s$  were determined on the bases of Eq. (1) and Fig. 1. The variables pertaining to concrete mixture design, environmental conditions and exposure time serve as inputs for the ML platform, whereas the corresponding values of  $C_s$  serve as the output. The concrete mixture design consists of eight variables representing the contents (units of kg/m<sup>3</sup>) of cement, fly ash, blast furnace slag, silica fume, superplasticizer, water, fine aggregate, and coarse aggregate. The environmental conditions are represented by two variables, that is, annual mean temperature (units of °C), and chloride concentration in seawater (units of g/L). Exposure time is one variable (units of annual, or a). An additional input – corresponding to environmental conditions – was added to represent the conditions to which concrete was exposed, that is, 0 for tidal zone, 1 for splash zone, and 2 for submerged zone. The distributions of the 13 attributes (i.e., 12 inputs and 1 output) in the dataset are summarized in Table 1.

### 3.2. Training and performance evaluation of machine learning models

For training and assessment of prediction performance of ML models,  $C_s$  dataset was randomly partitioned into two sets: a training set and a testing set. 75% of data-records of the parent dataset were used for training of the ML models (i.e., for fine-tuning, and, ultimately,

**Table 1**  
Statistical parameters pertaining to each of the attributes of  $C_s$  dataset.

|        | Attribute                                            | Unit              | Min.   | Max.    | Mean   | Std. Dev. |
|--------|------------------------------------------------------|-------------------|--------|---------|--------|-----------|
| Inputs | Cement                                               | kg/m <sup>3</sup> | 110.00 | 519.00  | 370.81 | 75.84     |
|        | Fly ash                                              | kg/m <sup>3</sup> | 0.00   | 239.00  | 33.88  | 59.77     |
|        | Blast furnace slag                                   | kg/m <sup>3</sup> | 0.00   | 292.50  | 11.23  | 44.77     |
|        | Silica fume                                          | kg/m <sup>3</sup> | 0.00   | 50.00   | 5.41   | 12.84     |
|        | Superplasticizer                                     | kg/m <sup>3</sup> | 0.00   | 10.20   | 1.46   | 1.98      |
|        | Water                                                | kg/m <sup>3</sup> | 38.50  | 311.00  | 187.90 | 44.20     |
|        | Fine aggregate                                       | kg/m <sup>3</sup> | 552.00 | 1232.00 | 768.34 | 120.04    |
|        | Coarse aggregate                                     | kg/m <sup>3</sup> | 410.00 | 1744.00 | 999.00 | 155.93    |
|        | Exposure time                                        | year              | 0.08   | 48.65   | 4.24   | 6.28      |
|        | Annual mean temperature (T)                          | °C                | 7.00   | 50.00   | 17.78  | 9.38      |
|        | Chloride content in seawater                         | g/L               | 13.00  | 21.45   | 19.25  | 3.55      |
|        | Exposure type (tidal = 0; splash = 1; submerged = 2) | –                 | –      | –       | –      | –         |
| Output | $C_s$                                                | % wt. concrete    | 0.14   | 13.58   | 3.67   | 2.09      |

finalizing, the optimum model parameters, assisted by the 10-fold cross-validation described in Section 2), and the remaining 25% were used for testing (i.e., for determination of cumulative error between predicted and actual values). Such split of 75%-to-25% between the training and test sets – or a ratio close to that – have been used in various past studies [66,81,82]. While the splitting was done randomly, special care was taken to guarantee that the training dataset was representative of the parent dataset. Towards this, it was ensured that the training dataset comprised of input attributes (i.e., concrete mixture design variables, environmental conditions, and exposure time) with widespread values encompassing the entire range between the two extrema in all the three exposure types.

For quantitative measurement of prediction performance of the ML models (against the test set), five different statistical parameters were used [83,84]. The parameters, essentially, estimate the cumulative error in predictions – of surface chloride concentration in the test dataset – with respect to the actual measurements. The statistical parameters are: Person correlation coefficient ( $r$ ), coefficient of determination ( $R^2$ ), mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean squared error (RMSE). The mathematical formulations to estimate these errors are shown in Eqs. (2)–(7); here,  $y'$  and  $y$  are predicted and measured values, and  $n$  is the total number of data-records in the test dataset.

$$r = \frac{n \sum y \cdot y' - (\sum y)(\sum y')}{\sqrt{n(\sum y^2) - (\sum y)^2} \sqrt{n(\sum y'^2) - (\sum y')^2}} \quad (2)$$

$$R^2 = \left[ \frac{n \sum y \cdot y' - (\sum y)(\sum y')}{\sqrt{n(\sum y^2) - (\sum y)^2} \sqrt{n(\sum y'^2) - (\sum y')^2}} \right]^2 \quad (3)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|y - y'|}{y} \quad (4)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y - y'| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |y - y'|^2} \quad (6)$$

$$CPI = \frac{1}{N} \sum_{j=1}^N \frac{P_j - P_{min,j}}{P_{max,j} - P_{min,j}} \quad (7)$$

The values of  $r$  would range from 0 to 1, wherein 0 represents the worst linear correlation and 1 represents the best linear correlation between the predicted values of model and the measured values. The values of  $R^2$  would range from 0 to 1 – the closer the value is to 1, the higher fitting optimization of the model is. The values MAE, MAPE and RMSE are used to evaluate model quality – the larger the value, the greater the difference between the predicted value and the measured value, that is, the worse the prediction of the model. To obtain a comprehensive measure of prediction performance of the ML models – and to compare them – the five statistical parameters described in Eqs. (2)–(6) were unified into a composite performance index (CPI, see Eq. (7)) [66,85]. In Eq. (7),  $N$  is the total number of performance measures (=5, as five statistical parameters were used in this study),  $P_j$  is the value of the  $j$ th statistical parameter, and  $P_{j, min}$  and  $P_{j, max}$  are the minimum and maximum values of the  $j$ th statistical parameter across the five values generated by the same number of ML models. Based on the formulation shown in Eq. (7), the values of CPI would range from 0 to 1, wherein 0 (or the lowest value) would represent the best ML model and 1 (or the maximum value) would represent the worst ML model in terms of overall prediction performance. In this study, the different ML models were ranked – from worst to best in terms of prediction performance – on the basis of their CPI values. The overall training and testing process of the ML models used in this study is described in Fig. 2.

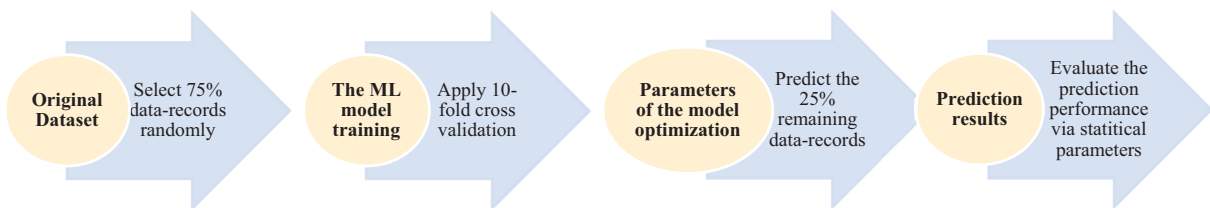
#### 4. Comparison and discussion of models

Based on the above discussions, the collected  $C_s$  data of marine concrete are used to train the six ML models and to assess their prediction performance. Their prediction performances are compared in this section. Furthermore, eight conventional regression-based quantitative models based on the same dataset of  $C_s$  are compared with the ML models to verify the advantages and practicability of the ML models.

##### 4.1. Comparison of ML models

As discussed previously, 75% of the collected data were used to train the six ML models, which were then used to predict  $C_s$  subjected to variations in relation to the 12 input variables. Predictions were then compared with the rest 25% database, as shown in Fig. 3, to determine the prediction performances of the models. For this purpose, the five statistical parameters (i.e.,  $r$ ,  $R^2$ , MAE, MAPE, and RMSE) and the composite performance index (CPI) were calculated according to Eqs. (2)–(7), and the results are shown in Table 2.

Overall, all ML models presented in this study were able to predict the  $C_s$  with reasonable accuracy, as shown in Fig. 3 and Table 2. This is evidenced by the relatively low RMSE (ranging between 0.11 and



**Fig. 2.** Flowchart for training and testing process of ML models.

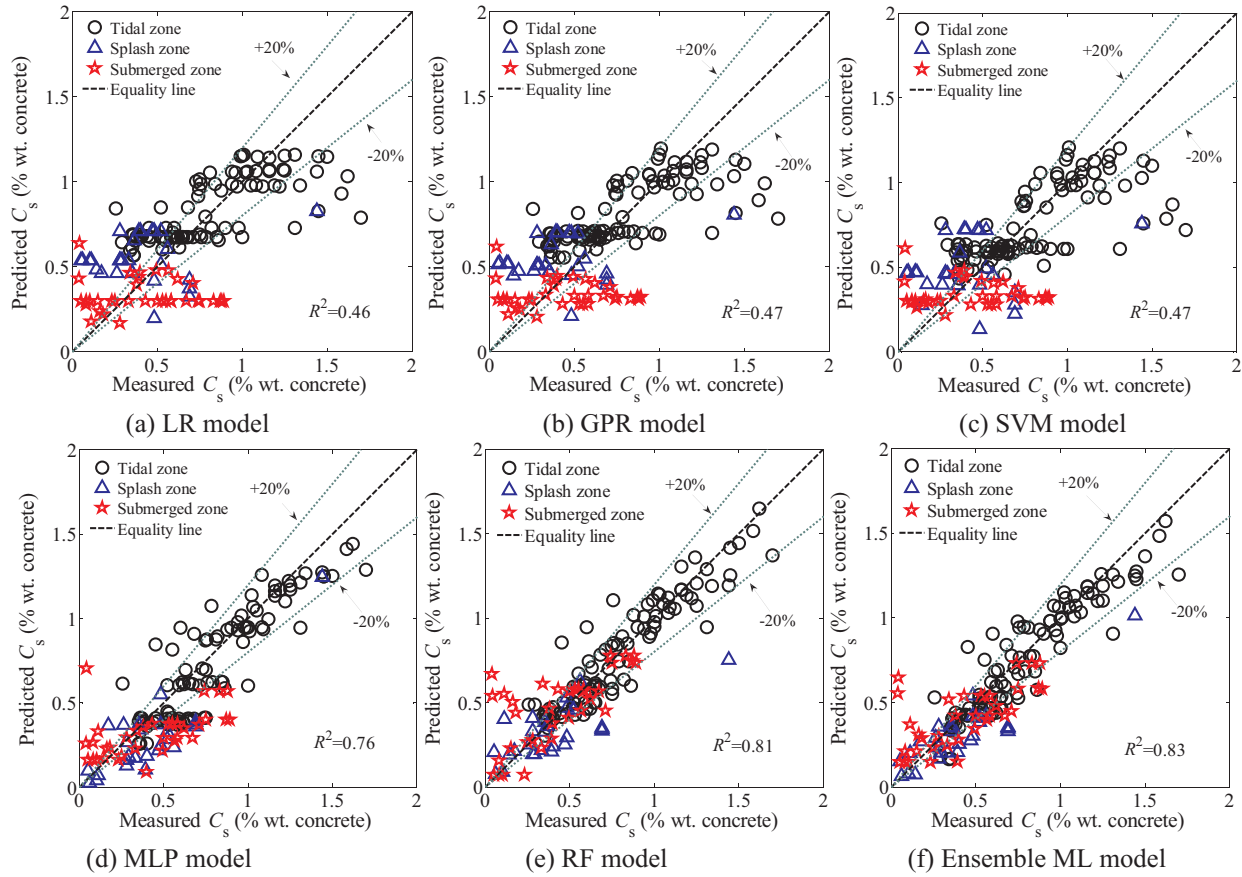


Fig. 3. Predictions of  $C_s$  made by ML models.

0.22% by weight of concrete) and relatively high  $R^2$  values (ranging between 0.46 and 0.83) of predictions made by the ML models, given the large size of the testing dataset. The best standalone model is RF (CPI is 0.02) and the worst is LR (CPI is 1). As can be seen in Fig. 3, only few predicted values of GPR fit within the  $\pm 20\%$  bound lines, but the majority of predicted values of RF fall within the  $\pm 20\%$  bound lines. The ensemble ML model achieved the most accurate outcomes of all ML models. It has the lowest CPI (i.e., 0.02), highest  $R^2$ , and the majority of its predictions fall within the  $\pm 20\%$  bound lines. These comparisons of prediction performances indicate that RF, MLP, and SVM, the top 3 standalone models, when combined within the ensemble ML model, can generally perform better than the best individual models for prediction of  $C_s$  of concrete. As can be seen from Fig. 3, the prediction performance of the ensemble ML model is the best in all ML models. We use this conclusion as the basis for next section, comparing the ensemble ML model with the conventional models for verifying the superiority of ML model.

#### 4.2. Comparison of ML with conventional quantitative models

In this section, eight conventional quantitative models of  $C_s$  are selected from the literature, which are compared with the ensemble ML model for verifying the superiority of machine learning. Information of the selected quantitative models are shown in Table 3.

It can be seen in Table 3 that some general quantitative models are applicable to all exposure types (i.e., splash, tidal, and submerged zones), while some were developed for specific exposure types (e.g., Chalee's model is for tidal zone only, and Yang's model is for submerged zone only). Based on relevant portion of the testing dataset, prediction results of the eight quantitative models and the ensemble ML model are compared with experimental results in Fig. 4. Note that in Fig. 4(g), Cai's model and Yang's model, which were developed by the same group of researchers, are merged to cover all the three exposure types and named Cai-Yang model. Since the unit of  $C_s$  calculated by most models in Table 3 is the percentage of chloride by weight of binder, the same unit (% wt. binder) has been used in Fig. 4. The  $C_s$  values in % wt. concrete were converted into values in % wt. binder by considering the mixture proportion of concrete. The  $C_s$  values in both % wt. binder and

Table 2  
Prediction performance of ML models.

| Statistical parameter<br>ML model | $r$  | $R^2$ | MAE (% wt. concrete) | MAPE (%) | RMSE (% wt. concrete) | CPI  |
|-----------------------------------|------|-------|----------------------|----------|-----------------------|------|
| LR                                | 0.68 | 0.46  | 0.22                 | 72.93    | 0.27                  | 1.00 |
| GPR                               | 0.69 | 0.47  | 0.21                 | 71.70    | 0.27                  | 0.97 |
| SVM                               | 0.68 | 0.47  | 0.20                 | 68.72    | 0.27                  | 0.94 |
| MLP-ANN                           | 0.87 | 0.76  | 0.16                 | 54.85    | 0.21                  | 0.35 |
| RF                                | 0.90 | 0.81  | 0.11                 | 37.27    | 0.16                  | 0.02 |
| Ensemble (RF + MLP + SVM)         | 0.91 | 0.83  | 0.12                 | 39.09    | 0.16                  | 0.02 |

**Table 3**  
Conventional quantitative models on surface chloride concentration of concrete.

| No. | Model name                | Applicable zone  | Equation of model                                                                                                                                                                     |
|-----|---------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1   | DuraCrete model [12]      | All              | $C_s = A_c(w/b)$ , where $A_c$ is correction factor of binder type (%); $w/b$ is water-to-binder ratio                                                                                |
| 2   | LNEC model [20,103]       | All              | $C_s = 2.5(w/b)k_T C_b$ , where $k_T$ is the coefficient that accounts for the concrete temperature; $C_b$ denotes the $C_s$ in the particular conditions of the Portuguese coast (%) |
| 3   | Song's model [38]         | All              | $C_s = 1.52\ln(3.77t + 1)$ , where $t$ is exposure time (year)                                                                                                                        |
| 4   | Chalee's model [16]       | Tidal            | $C_s = [-0.379(w/b) + 2.064]\ln(t) + [4.078(w/b) + 1.011]$                                                                                                                            |
| 5   | Petcherdchoo's model [42] | Tidal            | $C_s = 10^{[0.814(w/b)-0.213]} + 2.11t^{0.5}$                                                                                                                                         |
| 6   | Costa's model [19]        | Tidal            | $C_s = 0.38t^{0.37}$                                                                                                                                                                  |
| 7   | Cai's model [37]          | Tidal and splash | $C_{s,ts} = 10.01A_c(w/b)(1 - e^{-0.96t})$                                                                                                                                            |
| 8   | Yang's model [18]         | Submerged zone   | $C_{s,sub} = 4.12A_c(w/b)C_{sw}(1 - e^{(-0.56t)})$ , where $C_{sw}$ is chloride concentration in seawater (%)                                                                         |

% wt. concrete have been included in the database provided in the Supplementary information 2.

To measure the effectiveness of different models, the Root-Mean-Square Error (RMSE),  $\delta$ , defined by Eq. (6), and the mean ratio between the predicted value and the experimental value,  $\mu$ , defined by Eq. (8) are used to quantitatively analyze the prediction accuracy of each model:

$$\mu = \frac{1}{n} \sum_{i=1}^n \frac{y'}{y} \quad (8)$$

The  $\delta$  is a reflection of the discretization of the predicted value of the model from the measured data – the smaller  $\delta$  value is, the smaller the discretization of the predicted results of the model is. The  $\mu$  reflects the degree to which the predicted value of  $C_s$  is close to measured value. The closer the  $\mu$  is to 1, the closer the predicted value is to the measured value, indicating the higher prediction accuracy of the model.

As can be seen from Fig. 4, the ensemble ML model has the highest fitting accuracy compared with conventional models of  $C_s$ . By comparing the eight conventional  $C_s$  models included in Table 3, as illustrated in Fig. 4, Cai-Yang model has the highest fitting accuracy according to the values of  $\delta$  (1.71%) and  $\mu$  (1.17). The fitting performances of the DuraCrete model and Song's model are secondary to that of Cai-Yang model. Chalee's model, Petcherdchoo's model, and Costa's model yield reasonable predictions, but can only be applied to the tidal zone. Meanwhile, Chalee's model predicts some negative values of  $C_s$  because of the adopted function type, as shown in Fig. 4(d), which are unrealistic. Such comprehensive comparison of conventional models highlights that a model adopting wider data source and considering more influencing factors can perform better in predicting  $C_s$ . It is important to emphasize that, in Fig. 4, the  $C_s$  database was not processed by any filtering, such as deleting specious data. As compared with the conventional quantitative models, the ensemble ML method is developed based on randomly selected, large-size database, and takes into account all of the 12 input variables (see Table 1). These are unique advantages of ML, as it is difficult, if not impossible, for any conventional models to exhaustively account all variables. Due to these merits, logically, the ML model should have superior prediction performance compared to conventional models. Indeed, this hypothesis has been tested in Fig. 4 and validated by the accurate predictions achieved by the ensemble ML method ( $\delta = 0.94\%$ ,  $\mu = 1.18$ ).

In the establishment of conventional quantitative  $C_s$  models, it is necessary to screen and analyze data as well as influential factors; judge and select dominant and secondary influential factors; and try to find the best correlations between  $C_s$  and individual influential factors through fittings [6,16,28,37]. A multi-factor quantitative  $C_s$  model can then be established, adjusted, and finalized. Nevertheless, the ensemble ML model does not necessitate these time-consuming efforts, and can be employed without the assumption – or, even without comprehensive understanding – of correlations between input variables and output. The ensemble ML model is significantly more time-efficient, yet far more accurate than conventional methods. Furthermore, variables of

the ML model used for  $C_s$  prediction are all material and environmental parameters, which are easier to obtain than some of the parameters of the convectional models, such as the correction factor of binder ( $A_c$ ). In addition, with time, as the database expands, the ML model is expected to progressively become more accurate, ultimately opening the possibility of its utilization for optimization purposes (e.g., optimization of concrete mixture design to enhance its durability).

#### 4.3. Application of the ensemble ML model

Previous sections have shown that the ensemble ML model is able to reliably predict chloride concentration on the concrete surface in relation to concrete mixture design and environmental characteristics. In this section, the training of the ensemble ML model, based on the entire database, is leveraged to make blind predictions of surface chloride concentrations of new mixture designs and environmental conditions. To this end, seven scenarios are considered to examine the evolution of  $C_s$  as affected by seven factors, that is, exposure condition (or action class) and six major mixture proportioning parameters, as shown in Table 4. In scenario I, the exposure condition (i.e., tidal, splash, and submerged zones) is the only variable, while all other attributes are kept unchanged. In scenario II, the effect of  $w/b$  (i.e., 0.35, 0.45 and 0.55) is examined. To do this, the aggregate structure and content as well as environmental attributes are kept constant, and the amounts of binders and water are adjusted to meet the pre-set  $w/b$  and total volume of concrete. Following protocols similar to those for scenario II, five other mixture proportioning parameters are examined: fly ash (FA) replacement level (0%, 15% and 35%, scenario III); ground granulated blast-furnace slag (GGBS) replacement level (0%, 20% and 40%, scenario IV); silica fume (SF) replacement level (0%, 2.5% and 7.5%, scenario V); binder content (300 kg/m<sup>3</sup>, 400 kg/m<sup>3</sup> and 500 kg/m<sup>3</sup>, scenario VI); and sand ratio (fine aggregate/total aggregate weight ratio, 0.4, 0.44 and 0.48, scenario VII). Each scenario comprises of three cases, at least one of which is from the literature so we can find field data to directly compare with the predictions.

Under these seven scenarios, the ensemble ML model is used to predict the values of  $C_s$  after different exposure times. The obtained results are shown in Fig. 5. As shown in Fig. 5(a), the same concrete at a specific location can establish the highest  $C_s$  when exposed to tidal zone, followed by splash zone, and then by submerged zone. This is consistent with our speculation that the submerged zone has the lowest level of  $C_s$  due to absence of capillary suction. Fig. 5(b) indicates that the higher the  $w/b$ , the higher the value of  $C_s$  after the same exposure time. The exposure condition and  $w/b$  appear to be the most significant factors affecting the evolution and magnitude of  $C_s$ . When the  $w/b$  is kept constant, as seen in Fig. 5(c), the replacement of cement by 15% fly ash leads to higher  $C_s$ , but a higher replacement of 35% results in lower  $C_s$ . This nonmonotonic variation cannot be explained based on current knowledge, since fly ash is a highly variable material and the chemical composition of fly ash is missing in most of the references. The variation is more likely to be caused by insufficient volume and

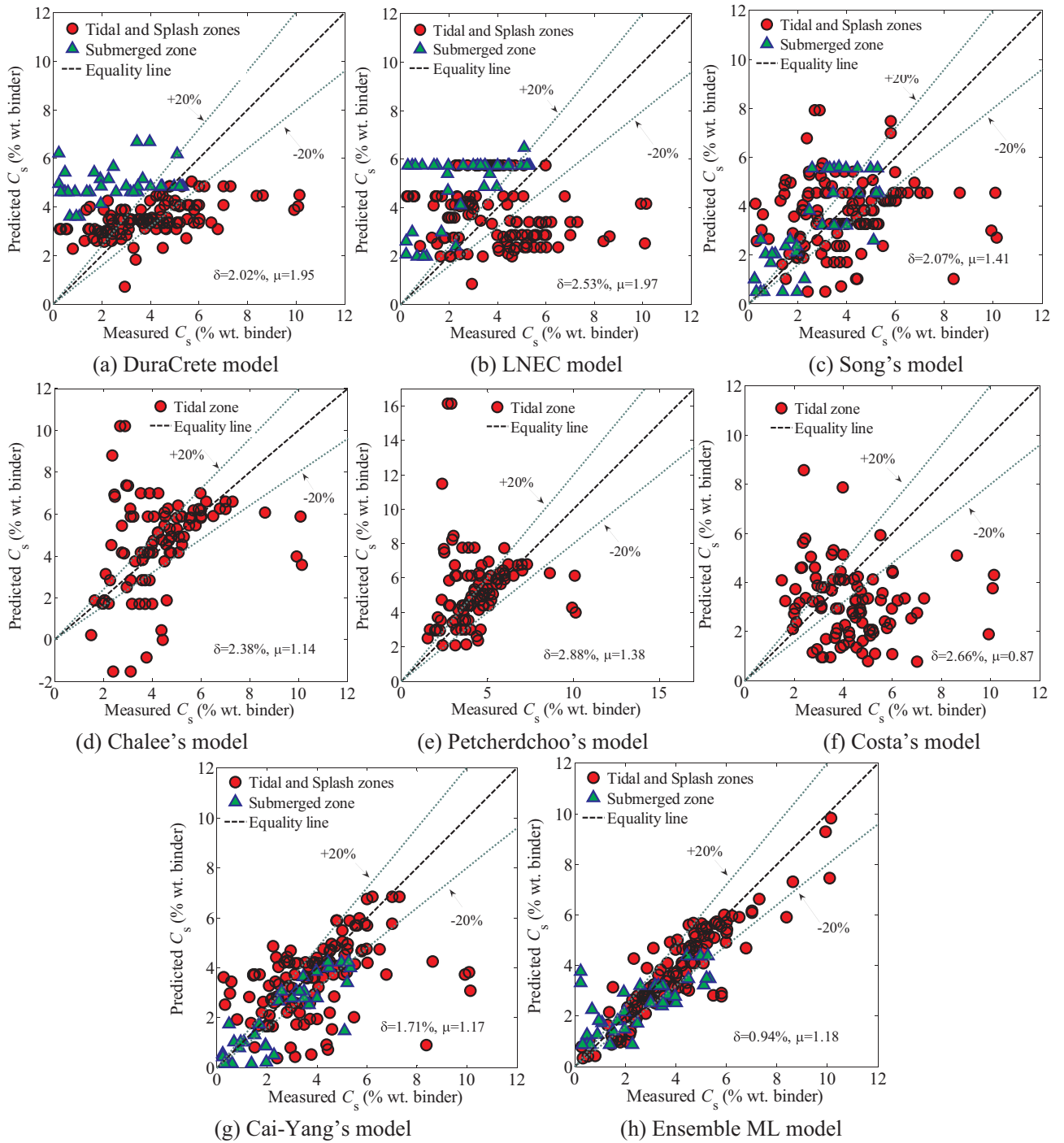


Fig. 4. Comparison of predicted results of different models based on experimental data.

diversity of the database, which may have resulted in inadequate training of the ML thereby resulting in some unreliable or unverifiable predictions. On the same note, in future studies, it will be important to consolidate larger database, and explicitly account for the composition of fly ash, in order to adequately train the ML model and enable reliable predictions. We would like to emphasize that majority of the ML-predicted trends revealed in scenarios I, II and III are consistent with the literature. Selected data from Chalee et al.'s work [16] has been plotted in Fig. 5(a) to (c) to enable direct comparison with the predictions. Fig. 5(d) shows that the incorporation of GGBS tends to lower the  $C_s$ , but this influence seems to be insignificant. Data at only one exposure time is available from reference [30] to compare with the predictions.

Since GGBS is also a highly variable material, the points that we made in the context of fly ash are also applicable to GGBS. As shown in Fig. 5(e), verified by limited data from reference [74], replacement of cement by silica fume tends to increase the  $C_s$ . Ghods et al. [70] and Farahani et al. [72] reported the same trends in various exposure conditions (i.e., tidal, splash, and submerged zones). This doesn't necessarily mean poorer performance, since the incorporation of silica fume can substantially decrease the diffusion coefficient of concrete [70]. As partially verified by data from reference [22], the effects of binder content and sand ratio on  $C_s$  are illustrated in Fig. 5(f) and (g). According to the trends shown in the figures, higher binder content and higher sand ratio tend to decrease and increase  $C_s$ , respectively. These

**Table 4**  
Combinations of the dataset attributes for blind predictions of  $C_s$ .

|              | Zone     | OPC, kg/<br>m <sup>3</sup> | FA, kg/<br>m <sup>3</sup> | GGBS, kg/<br>m <sup>3</sup> | SF, kg/<br>m <sup>3</sup> | SuperP, kg/<br>m <sup>3</sup> | Water, kg/<br>m <sup>3</sup> | Fine Agg,<br>kg/m <sup>3</sup> | Coarse Agg,<br>kg/m <sup>3</sup> | Mean<br>Temp,<br>°C | [Cl <sup>-</sup> ] in seawater,<br>g/L | w/b  |
|--------------|----------|----------------------------|---------------------------|-----------------------------|---------------------------|-------------------------------|------------------------------|--------------------------------|----------------------------------|---------------------|----------------------------------------|------|
| Scenario I   | Tidal    | 406                        | 72                        | 0                           | 0                         | 0                             | 215                          | 639                            | 1024                             | 30                  | 17                                     | 0.45 |
|              | Splash   | 406                        | 72                        | 0                           | 0                         | 0                             | 215                          | 639                            | 1024                             | 30                  | 17                                     | 0.45 |
|              | Submerge | 406                        | 72                        | 0                           | 0                         | 0                             | 215                          | 639                            | 1024                             | 30                  | 17                                     | 0.45 |
| Scenario II  | Tidal    | 466                        | 82                        | 0                           | 0                         | 0                             | 192                          | 639                            | 1024                             | 30                  | 17                                     | 0.35 |
|              | Tidal    | 406                        | 72                        | 0                           | 0                         | 0                             | 215                          | 639                            | 1024                             | 30                  | 17                                     | 0.45 |
|              | Tidal    | 360                        | 64                        | 0                           | 0                         | 0                             | 233                          | 639                            | 1024                             | 30                  | 17                                     | 0.55 |
| Scenario III | Tidal    | 478                        | 0                         | 0                           | 0                         | 0                             | 215                          | 639                            | 1024                             | 30                  | 17                                     | 0.45 |
|              | Tidal    | 406                        | 72                        | 0                           | 0                         | 0                             | 215                          | 639                            | 1024                             | 30                  | 17                                     | 0.45 |
|              | Tidal    | 311                        | 167                       | 0                           | 0                         | 0                             | 215                          | 639                            | 1024                             | 30                  | 17                                     | 0.45 |
| Scenario IV  | Tidal    | 430                        | 0                         | 0                           | 0                         | 1                             | 163                          | 800                            | 1000                             | 14                  | 19                                     | 0.38 |
|              | Tidal    | 335                        | 0                         | 84                          | 0                         | 1                             | 159                          | 800                            | 1000                             | 14                  | 19                                     | 0.38 |
|              | Tidal    | 240                        | 0                         | 165                         | 0                         | 1                             | 154                          | 800                            | 1000                             | 14                  | 19                                     | 0.38 |
| Scenario V   | Tidal    | 410                        | 0                         | 0                           | 0                         | 3.2                           | 164                          | 832                            | 1017                             | 27.5                | 23.96                                  | 0.4  |
|              | Tidal    | 400                        | 0                         | 0                           | 10                        | 3.2                           | 164                          | 832                            | 1017                             | 27.5                | 23.96                                  | 0.4  |
|              | Tidal    | 370                        | 0                         | 0                           | 30                        | 3.2                           | 160                          | 832                            | 1017                             | 27.5                | 23.96                                  | 0.4  |
| Scenario VI  | Submerge | 300                        | 0                         | 0                           | 0                         | 1                             | 135                          | 888                            | 1132                             | 7                   | 19                                     | 0.45 |
|              | Submerge | 400                        | 0                         | 0                           | 0                         | 1                             | 180                          | 800                            | 1020                             | 7                   | 19                                     | 0.45 |
|              | Submerge | 500                        | 0                         | 0                           | 0                         | 1                             | 225                          | 710                            | 905                              | 7                   | 19                                     | 0.45 |
| Scenario VII | Submerge | 400                        | 0                         | 0                           | 0                         | 1                             | 180                          | 728                            | 1092                             | 7                   | 19                                     | 0.45 |
|              | Submerge | 400                        | 0                         | 0                           | 0                         | 1                             | 180                          | 800                            | 1020                             | 7                   | 19                                     | 0.45 |
|              | Submerge | 400                        | 0                         | 0                           | 0                         | 1                             | 180                          | 874                            | 946                              | 7                   | 19                                     | 0.45 |

trends seem to be consistent with the microstructural features of concrete – less binder or more sand can lead to more interconnected interfacial transition zones (ITZs) [86], and ITZs facilitate ingress of ions in the surface layer of concrete. However, the differences in  $C_s$  caused by these two factors tend to diminish and ultimately disappear following the increasing of exposure time. Both scenarios VI and VII are in the submerged zone, wherein the effects of binder content and sand ratio appear to be minuscule. Nevertheless, these effects could be more significant in splash zone and tidal zone wherein capillary actions, which play a significant role in transport of species within the surface layer of concrete, feature predominantly.

## 5. Conclusions

Based on extensive mining of data from the literature, 642 data-records of field exposure data of surface chloride concentration  $C_s$  in marine concrete are collected, which cover splash, tidal, and submerged zones in various areas in the world. The database is used to rigorously train, and the test, machine learning (ML) models to enable accurate predictions of  $C_s$ . Six standalone, that is, linear regression (LR), Gaussian process regression (GPR), support vector machine (SVM), multilayer perceptron artificial neural network (MLP-ANN), random Forests (RF), and an ensemble weighted-voting based ML model, are introduced in the paper. The accuracy of these ML models in predicting  $C_s$  are compared and analyzed. The best performing ML model is selected to compare with eight general and traditional quantitative  $C_s$  models, so as to verify the advantages and practicability of ML. The verified model is also employed to predict evolutions of  $C_s$  of mix proportions not included in the database, so as to examine the effects of seven factors on  $C_s$ . Through this study, the following conclusions can be drawn:

- (1) By meta-heuristically combining the predictions of multiple standalone ML models, an ensemble ML model produces superior prediction performance compared to anyone of the standalone ML models.
- (2) The process of establishment of conventional quantitative  $C_s$  model is complex, yet tedious. A model that adopts more diverse data source during establishment and considers more influencing factors may perform better in predicting  $C_s$ . As evaluated based on the testing dataset (randomly selected 25% of the database), Cai-Yang

model is the best-performing one in all the conventional models included in this paper.

- (3) The ensemble ML model can easily consider multiple input variables (12 in this study), and, thus, has the potential to perform better than simple conventional models. In general, the larger the database, the more accurate the predictions produced by the ensemble ML model.
- (4) The exposure condition (i.e., tidal, splash and submerged zones) and w/b appear to be the most significant factors affecting  $C_s$ . When other factors are fixed, concrete exposed to tidal zone tends to have the highest  $C_s$ , followed by splash zone and submerged zone. A higher w/b leads to larger  $C_s$ . In addition, a higher binder content and a lower sand ratio could retard the establishment of  $C_s$ .

To sum up, an ensemble ML model is successfully applied in the present study to predict surface chloride concentration of marine concrete. It appears to be an accurate, yet computationally-efficient method. However, the model can be further improved in the future, for example, by extending the database and considering more compositional factors. The database will have to be extended by adding expert-endorsed data to cover much broader diversities in mixture proportions and service environments. Apart from weight-based mixture proportioning parameters, the chemical compositions of the binder, especially supplementary cementitious materials, may need to be considered in future ML models. After all, a supplementary cementitious material (e.g., fly ash) could represent a class of highly variable materials, instead of one unique material. Ignoring their compositional features could lead to large errors that counteract the advantages of ML models. In the end, it is worth noting that the ensemble ML model has a wide adaptability. It can readily be adapted to predict other durability-related properties of concrete (e.g., diffusion coefficient) simultaneously given reliable and large enough relevant database. It is, therefore, expected that well trained-and-validated ML models can aid in durability design or service life prediction of concrete structure located in different service environments.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cemconres.2020.106164>.

## CRediT authorship contribution statement

**Rong Cai:** Conceptualization, Data curation, Writing - original

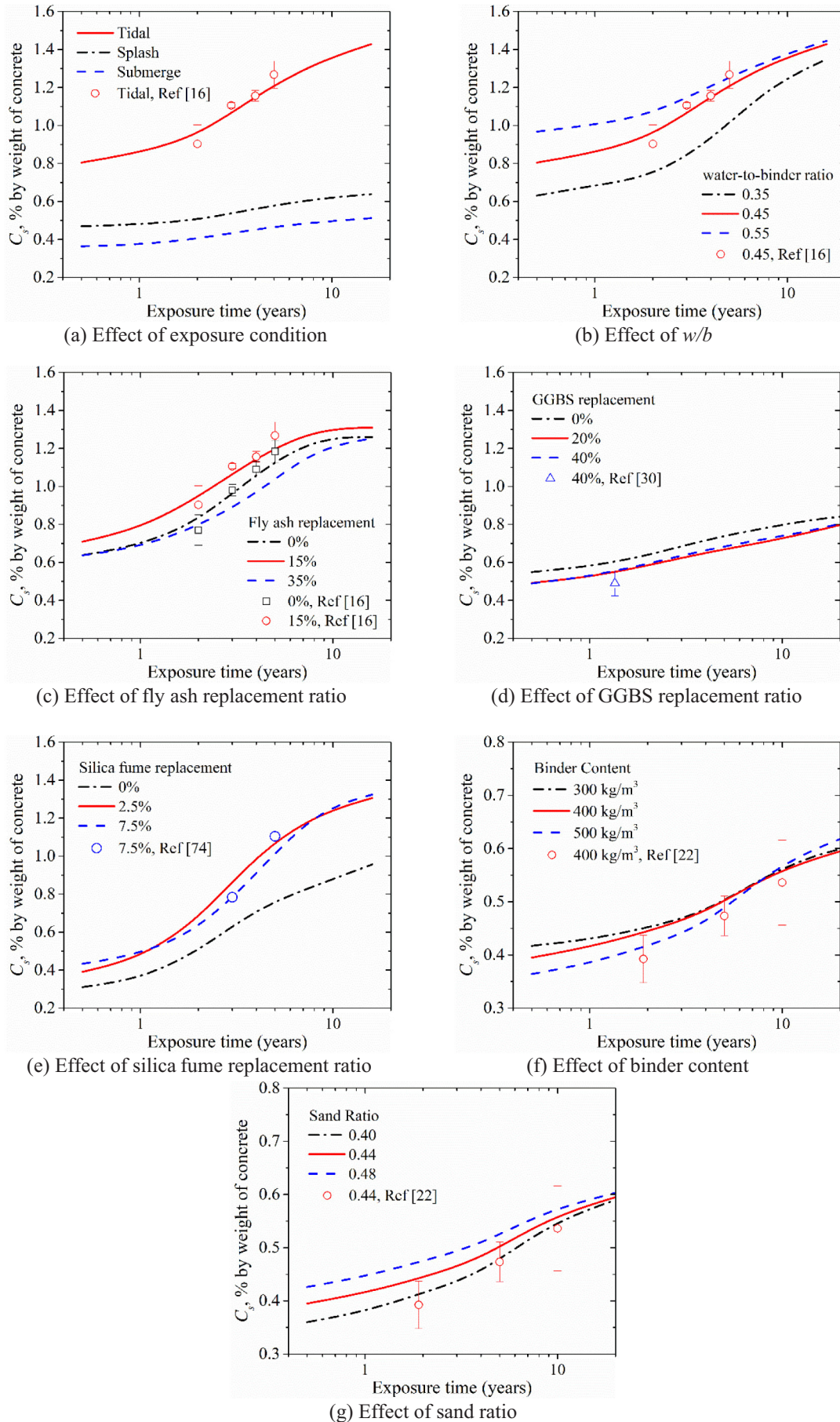


Fig. 5. Evolution of  $C_s$  as affected by seven factors predicted by the ensemble ML model.

draft. **Taihao Han**: Methodology, Software, Writing - original draft. **Wenyu Liao**: Data curation, Writing - original draft. **Jie Huang**: Investigation, Software. **Dawang Li**: Data curation. **Aditya Kumar**: Methodology, Software, Writing - review & editing, Funding acquisition. **Hongyan Ma**: Conceptualization, Supervision, Writing - review & editing, Funding acquisition.

## Declaration of competing interest

The authors declare no conflict of interests.

## Acknowledgements

Financial support from the National Science Foundation (under CMMI-1661609 and 1761697) and the Leonard Wood Institute (LWI) is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of the National Science Foundation. The first author would like to gratefully acknowledge the financial support from the Education Department of Guangxi Zhuang Autonomous Region, Guangxi Key Laboratory of Disaster Prevention and Engineering Safety and Key Laboratory of Disaster Prevention and Structural Safety of Ministry of Education, Guangxi University, P. R. China. The fourth author would like to acknowledge the support from the National Natural Science Foundation of China under 51520105012. In addition, the support from the Advanced Materials for Sustainable Infrastructure signature area at Missouri S&T is also acknowledged. Computational tasks were conducted in the Materials Research Center and Department of Materials Science and Engineering at Missouri S&T. The authors gratefully acknowledge the financial support that has made these laboratories and their operations possible.

## References

- [1] O. Gjorv, *Durability Design of Concrete Structures in Severe Environments*, Taylor and Francis, Norway, 2009.
- [2] C.A. Apostolopoulos, S. Demis, V.G. Papadakis, Chloride-induced corrosion of steel reinforcement – mechanical performance and pit depth analysis, *Constr. Build. Mater.* 38 (2013) 139–146, <https://doi.org/10.1016/j.conbuildmat.2012.07.087>.
- [3] C.E.T. Balestra, M.G. Lima, A.R. Silva, R.A. Medeiros-Junior, Corrosion degree effect on nominal and effective strengths of naturally corroded reinforcement, *J. Mater. Civ. Eng.* 28 (2016) 04016103, [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0001599](https://doi.org/10.1061/(ASCE)MT.1943-5533.0001599).
- [4] G.R. Meira, C. Andrade, C. Alonso, I.J. Padaratz, J.C. Borba, Modelling sea-salt transport and deposition in marine atmosphere zone – a tool for corrosion studies, *Corros. Sci.* 50 (2008) 2724–2731, <https://doi.org/10.1016/j.corsci.2008.06.028>.
- [5] C.E.T. Balestra, T.A. Reichert, G. Savaris, Contribution for durability studies based on chloride profiles analysis of real marine structures in different marine aggressive zones, *Constr. Build. Mater.* 206 (2019) 140–150, <https://doi.org/10.1016/j.conbuildmat.2019.02.067>.
- [6] L.F. Yang, R. Cai, B. Yu, Investigation of computational model for surface chloride concentration of concrete in marine atmosphere zone, *Ocean Eng.* 138 (2017) 105–111, <https://doi.org/10.1016/j.oceaneng.2017.04.024>.
- [7] F. Rosquoët, S. Bonnet, F. Schoefs, A. Khelidj, Chloride propagation in concrete harbour, 2nd International Symposium on Advances in Concrete through Science and Engineering, Québec, Canada, 2006 <https://hal.archives-ouvertes.fr/hal-01007980>, Accessed date: 28 January 2020.
- [8] W.Z. Taffesea, E. Sistonon, Service life prediction of repaired structures using concrete recasting method: state-of-the-art, *Procedia Engineering* 57 (2013) 1138–1144, <https://doi.org/10.1016/j.proeng.2013.04.143>.
- [9] Y. Zhang, Optimal sustainable life cycle maintenance strategies for port infrastructures, *J. Clean. Prod.* 142 (2017) 17–1709, <https://doi.org/10.1016/j.jclepro.2016.11.120>.
- [10] A. Costa, J. Appleton, Chloride penetration into concrete in marine environment—part II: prediction of long term chloride penetration, *Mat. Struct.* 32 (1999) 354–359, <https://doi.org/10.1007/BF02479627>.
- [11] CEN, *Concrete-Part 1: Specification, Performance, Production and Conformity*, EN 206-1, C. E. De Normalisation, Brussels, 2000.
- [12] *DuraCrete, Final Technical Report-General Guidelines for Durability Design and Redesign*, Project Document BE95-1347/R17, Denmark, (2000).
- [13] W. Chalee, C. Jaturapitakkul, Effects of W/B ratios and fly ash finenesses on chloride diffusion coefficient of concrete in marine environment, *Mater. Struct.* 42 (2009) 505–514, <https://doi.org/10.1617/s11527-008-9398-2>.
- [14] M. Collepardi, A. Marcialis, R. Turriziani, Penetration of chloride ions into cement pastes and concretes, *J. American Ceramic Society* 55 (1972) 534–535, <https://doi.org/10.1111/j.1151-2916.1972.tb13424.x>.
- [15] I. Stipanovic Oslakovic, D. Bjegovic, D. Mikulic, Evaluation of service life design models on concrete structures exposed to marine environment, *Mater. Struct.* 43 (2010) 1397–1412, <https://doi.org/10.1617/s11527-010-9590-z>.
- [16] W. Chalee, C. Jaturapitakkul, P. Chindaprasit, Predicting the chloride penetration of fly ash concrete in seawater, *Mar. Struct.* 22 (2009) 341–353, <https://doi.org/10.1016/j.marstruc.2008.12.001>.
- [17] L. Yang, R. Cai, B. Yu, Computing model for surface chloride concentration of concrete in marine tidal and splash zones, *The Ocean Engineering* 32 (2014) 25–33, <https://doi.org/10.16483/j.issn.1005-9865.2014.05.003>.
- [18] L. Yang, R. Cai, B. Yu, Modeling of environmental action for submerged marine concrete in terms of surface chloride concentration, *Struct. Concr.* 19 (2018) 1512–1520, <https://doi.org/10.1002/suco.201800072>.
- [19] A. Costa, J. Appleton, Chloride penetration into concrete in marine environment—part I: main parameters affecting chloride penetration, *Mat. Struct.* 32 (1999) 252–259, <https://doi.org/10.1007/BF02479594>.
- [20] P.F. Marques, A. Costa, F. Lanata, Service life of RC structures: chloride induced corrosion: prescriptive versus performance-based methodologies, *Mater. Struct.* 45 (2012) 277–296, <https://doi.org/10.1617/s11527-011-9765-2>.
- [21] S. Nanukuttan, L. Basheer, W.J. McCarter, D. Robinson, P.A.M. Basheer, Full-scale marine exposure tests on treated and untreated concretes—initial 7-year results, *ACI Mater. J.* 105 (2008) 81–87.
- [22] G. Markeset, O. Skjolsvold, Time dependent chloride diffusion coefficient - field studies of concrete exposed to marine environment in Norway, *Proceedings of the 2nd International Symposium on Service Life Design for Infrastructure*, 2010.
- [23] M. Safedian, A.A. Ramezaniapour, Prediction of RC structure service life from field long term chloride diffusion, *Comput. Concr.* 15 (2015) 589–606, <https://doi.org/10.12989/CAC.2015.15.4.589>.
- [24] L. Tang, Chloride ingress in concrete exposed to marine environment: field data up to 10 years exposure, *Sveriges provnings-och forskningsinstitut (SP)*, <https://books.google.com/books?id=TrUCNQAAACAAJ>, (2003).
- [25] G.R. Meira, C. Andrade, C. Alonso, J.C. Borba, M. Padilha, Durability of concrete structures in marine atmosphere zones – the use of chloride deposition rate on the wet candle as an environmental indicator, *Cem. Concr. Compos.* 32 (2010) 427–435, <https://doi.org/10.1016/j.cemconcomp.2010.03.002>.
- [26] A. Lindvall, Chloride ingress data from field and laboratory exposure – influence of salinity and temperature, *Cem. Concr. Compos.* 29 (2007) 88–93, <https://doi.org/10.1016/j.cemconcomp.2006.08.004>.
- [27] G.R. Meira, C. Andrade, C. Alonso, I.J. Padaratz, J.C. Borba Jr, Salinity of marine aerosols in a Brazilian coastal area—influence of wind regime, *Atmos. Environ.* 41 (2007) 8431–8441, <https://doi.org/10.1016/j.atmosenv.2007.07.004>.
- [28] G.R. Meira, C. Andrade, I.J. Padaratz, C. Alonso, J.C. Borba, Chloride penetration into concrete structures in the marine atmosphere zone – relationship between deposition of chlorides on the wet candle and chlorides accumulated into concrete, *Cem. Concr. Compos.* 29 (2007) 667–676, <https://doi.org/10.1016/j.cemconcomp.2007.05.009>.
- [29] H.-W. Song, C.-H. Lee, K.Y. Ann, Factors influencing chloride transport in concrete structures exposed to marine environments, *Cem. Concr. Compos.* 30 (2008) 113–121, <https://doi.org/10.1016/j.cemconcomp.2007.09.005>.
- [30] S.-W. Pack, M.-S. Jung, H.-W. Song, S.-H. Kim, K.Y. Ann, Prediction of time dependent chloride transport in concrete structures exposed to a marine environment, *Cem. Concr. Res.* 40 (2010) 302–312, <https://doi.org/10.1016/j.cemconres.2009.09.023>.
- [31] X. Wang, W. Jin, Chloride penetration in wet-dry cycling zone of harbor concrete structure, *The Ocean Engineering* 28 (2010) 97–104, <https://doi.org/10.16483/j.issn.1005-9865.2010.04.020>.
- [32] Z. Jin, W. Xue, J. Chen, Effecting coefficients for concrete structure durability design index, *Journal of Building Structures* 32 (2011) 86–97, <https://doi.org/10.3969/j.issn.1002-3550.2017.05.014>.
- [33] W. Xue, W.L. Jin, H. Yokota, Chloride penetration under co-effects of initial curing and exposure conditions, *Journal of Zhejiang University (Engineering Science)* 45 (2011) 1416–1422, <https://doi.org/10.3785/j.issn.1008-973X.2011.08.015>.
- [34] Q. Li, K. Li, X. Zhou, Q. Zhang, Z. Fan, Model-based durability design of concrete structures in Hong Kong-Zhuhai-Macao sea link project, *Struct. Saf.* 53 (2015) 1–12, <https://doi.org/10.1016/j.strusafe.2014.11.002>.
- [35] S.N. Wang, K.F. Li, Z.H. Fan, Q.K. Su, J.B. Xiong, Durability strategy for main concrete structure of Hong Kong-Zhuhai-Macao bridge with designed service life of 120 years, *Port & Waterway Engineering* 501 (2015) 78–92, <https://doi.org/10.16233/j.cnki.issn1002-4972.2015.03.020>.
- [36] L. Pang, Q. Li, Service life prediction of RC structures in marine environment using long term chloride ingress data: comparison between exposure trials and real structure surveys, *Constr. Build. Mater.* 113 (2016) 979–987, <https://doi.org/10.1016/j.conbuildmat.2016.03.156>.
- [37] R. Cai, L.F. Yang, B. Yu, Improved model for surface chloride concentration of concrete in marine tidal and splash zones, *Journal of Civil and Environmental Engineering* 41 (2019) 122–129, <https://doi.org/10.11835/j.issn.2096-6717.2019.079>.
- [38] H.-W. Song, H.-B. Shim, A. Petcherdchoo, S.-K. Park, Service life prediction of repaired concrete structures under chloride environment using finite difference method, *Cem. Concr. Compos.* 31 (2009) 120–127, <https://doi.org/10.1016/j.cemconcomp.2008.11.002>.
- [39] S.H. Lin, Chloride diffusion in a porous concrete slab, *Corrosion* 46 (1990) 964–967, <https://doi.org/10.5006/1.3585052>.
- [40] P. Arora, B.N. Popov, B. Haran, M. Ramasubramanian, S. Popova, R.E. White, Corrosion initiation time of steel reinforcement in a chloride environment—a one dimensional solution, *Corros. Sci.* 39 (1997) 739–759, <https://doi.org/10.1016/>

- S0010-938X(96)00163-1.
- [41] M.K. Kassar, M. Ghosn, Chloride-induced corrosion of reinforced concrete bridge decks, *Cem. Concr. Res.* 32 (2002) 139–143, [https://doi.org/10.1016/S0008-8846\(01\)00644-5](https://doi.org/10.1016/S0008-8846(01)00644-5).
  - [42] A. Petcherdchoo, Time dependent models of apparent diffusion coefficient and surface chloride for chloride transport in fly ash concrete, *Constr. Build. Mater.* 38 (2013) 497–507, <https://doi.org/10.1016/j.conbuildmat.2012.08.041>.
  - [43] A. Goel, M. Pal, Application of support vector machines in scour prediction on grade-control structures, *Eng. Appl. Artif. Intell.* 22 (2009) 216–223, <https://doi.org/10.1016/j.engappai.2008.05.008>.
  - [44] Y.F. Wen, C.Z. Cai, X.H. Liu, J.F. Pei, X.J. Zhu, T.T. Xiao, Corrosion rate prediction of 3C steel under different seawater environment by using support vector regression, *Corros. Sci.* 51 (2009) 349–355, <https://doi.org/10.1016/j.corsci.2008.10.038>.
  - [45] J.-S. Chou, N.-T. Ngo, W.K. Chong, The use of artificial intelligence combiners for modeling steel pitting risk and corrosion rate, *Eng. Appl. Artif. Intell.* 65 (2017) 471–483, <https://doi.org/10.1016/j.engappai.2016.09.008>.
  - [46] L. Marques Alves, R. Almeida Cotta, P. Marques Ciarelli, Identification of types of corrosion through electrochemical noise using machine learning techniques, *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods, SCITEPRESS - Science and Technology Publications, Porto, Portugal, 2017*, pp. 332–340, <https://doi.org/10.5220/0006122403320340>.
  - [47] W.Z. Taffese, E. Sistonon, Machine learning for durability and service-life assessment of reinforced concrete structures: recent advances and future directions, *Autom. Constr.* 77 (2017) 1–14, <https://doi.org/10.1016/j.autcon.2017.01.016>.
  - [48] 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>.
  - [49] M. Nikoo, Ł. Sadowski, M. Nikoo, Prediction of the corrosion current density in reinforced concrete using a self-organizing feature map, *Coatings* 7 (2017) 160, <https://doi.org/10.3390/coatings7100160>.
  - [50] G. Morcous, Z. Lounis, Prediction of onset of corrosion in concrete bridge decks using neural networks and case-based reasoning, *Computer-Aided Civil and Infrastructure Engineering* 20 (2005) 108–117, <https://doi.org/10.1111/j.1467-8667.2005.00380.x>.
  - [51] I.H. Witten, E. Frank, M.A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed, Morgan Kaufmann, Burlington, MA, 2011.
  - [52] J.A. Nelder, R. Mead, A simplex method for function minimization, *Comput. J.* 7 (1965) 308–313, <https://doi.org/10.1093/comjnl/7.4.308>.
  - [53] 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.
  - [54] E. Schulz, M. Speekenbrink, A. Krause, A tutorial on Gaussian process regression: modelling, exploring, and exploiting functions, *BioRxiv* (2017), <https://doi.org/10.1101/095190>.
  - [55] R.J. Schalkoff, *Artificial Neural Networks*, McGraw-Hill New York, 1997.
  - [56] 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](https://doi.org/10.1016/S1352-2310(97)00447-0).
  - [57] T. Hegazy, P. Fazio, O. Moselhi, Developing practical neural network applications using back-propagation, *Computer-Aided Civil and Infrastructure Engineering* 9 (1994) 145–159, <https://doi.org/10.1111/j.1467-8667.1994.tb00369.x>.
  - [58] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, *Stat. Comput.* 14 (2004) 199–222, <https://doi.org/10.1023/B:STCO.0000035301.49549.88>.
  - [59] Vladimir Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed., Springer-Verlag, New York, 2000.
  - [60] C. Strobl, A.-L. Boulesteix, A. Zeileis, T. Hothorn, Bias in random forest variable importance measures: illustrations, sources and a solution, *BMC Bioinformatics* 8 (2007) 25, <https://doi.org/10.1186/1471-2105-8-25>.
  - [61] K.J. Archer, R.V. Kimes, Empirical characterization of random forest variable importance measures, *Computational Statistics & Data Analysis* 52 (2008) 2249–2260, <https://doi.org/10.1016/j.csda.2007.08.015>.
  - [62] L. Breiman, Bagging predictors, *Mach. Learn.* 24 (1996) 123–140, <https://doi.org/10.1007/BF00058655>.
  - [63] Y. Liu, T. Zhao, W. Ju, S. Shi, Materials discovery and design using machine learning, *J. Mater.* 3 (2017) 159–177, <https://doi.org/10.1016/j.jmat.2017.08.002>.
  - [64] R. Polikar, *Ensemble learning, Ensemble Machine Learning*, Springer, 2012, pp. 1–34.
  - [65] A.K. Tanwani, J. Afridi, M.Z. Shafiq, M. Farooq, Guidelines to select machine learning scheme for classification of biomedical datasets, *European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*, 2009, pp. 128–139, [https://doi.org/10.1007/978-3-642-01184-9\\_12](https://doi.org/10.1007/978-3-642-01184-9_12).
  - [66] 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>.
  - [67] T.G. Dietterich, *Ensemble methods in machine learning, International Workshop on Multiple Classifier Systems*, Springer, 2000, pp. 1–15.
  - [68] C. Schaffer, Selecting a classification method by cross-validation, *Mach. Learn.* 13 (1993) 135–143, <https://doi.org/10.1007/BF00993106>.
  - [69] B.L. Zhang, S.S. Wei, Decade Exposure Test of Harbor Reinforced Concrete for 10 Years, 98 South China Harbour Engineering, 1998, pp. 32–40.
  - [70] P. Ghods, M. Chini, R. Alizadeh, M. Hoseini, M. Shekarchi, A. Ramezaniapour, *The Effect of Different Exposure Conditions on the Chloride Diffusion into Concrete in the Persian Gulf Region*, Iran (2005).
  - [71] M. Valipour, F. Pargar, M. Shekarchi, S. Khani, M. Moradian, In situ study of chloride ingress in concretes containing natural zeolite, metakaolin and silica fume exposed to various exposure conditions in a harsh marine environment, *Constr. Build. Mater.* 46 (2013) 63–70, <https://doi.org/10.1016/j.conbuildmat.2013.03.026>.
  - [72] A. Farahani, H. Taghaddos, M. Shekarchi, Prediction of long-term chloride diffusion in silica fume concrete in a marine environment, *Cem. Concr. Compos.* 59 (2015) 10–17, <https://doi.org/10.1016/j.cemconcomp.2015.03.006>.
  - [73] Y. Gao, J. Zhang, S. Zhang, Y. Zhang, Probability distribution of convection zone depth of chloride in concrete in a marine tidal environment, *Constr. Build. Mater.* 140 (2017) 485–495, <https://doi.org/10.1016/j.conbuildmat.2017.02.134>.
  - [74] M. Khanzadeh Moradillo, S. Sadati, M. Shekarchi, Quantifying maximum phenomenon in chloride ion profiles and its influence on service-life prediction of concrete structures exposed to seawater tidal zone – a field oriented study, *Constr. Build. Mater.* 180 (2018) 109–116, <https://doi.org/10.1016/j.conbuildmat.2018.05.284>.
  - [75] Y. Wang, L. Wu, Y. Wang, Q. Li, Z. Xiao, Prediction model of long-term chloride diffusion into plain concrete considering the effect of the heterogeneity of materials exposed to marine tidal zone, *Constr. Build. Mater.* 159 (2018) 297–315, <https://doi.org/10.1016/j.conbuildmat.2017.10.083>.
  - [76] Y. Zhang, Y. Zhang, J. Huang, H. Zhuang, J. Zhang, Time dependence and similarity analysis of peak value of chloride concentration of concrete under the simulated chloride environment, *Constr. Build. Mater.* 181 (2018) 609–617, <https://doi.org/10.1016/j.conbuildmat.2018.06.030>.
  - [77] M. Safedian, A.A. Ramezaniapour, Assessment of service life models for determination of chloride penetration into silica fume concrete in the severe marine environmental condition, *Constr. Build. Mater.* 48 (2013) 287–294, <https://doi.org/10.1016/j.conbuildmat.2013.07.006>.
  - [78] H. Xue, Z.Q. Jin, X.J. Wang, Chloride ion penetration into concrete exposed to marine environment for a long period, *The Ocean Engineering* 33 (2015) 60–65, <https://doi.org/10.16483/j.issn.1005-9865.2015.05.008>.
  - [79] R. Alizadeh, P. Ghods, M. Chini, M. Hoseini, M. Ghalibafian, M. Shekarchi, Effect of curing conditions on the service life design of RC structures in the Persian Gulf Region, *J. Mater. Civ. Eng.* 20 (2008) 2–8, [https://doi.org/10.1061/\(ASCE\)0899-1561\(2008\)20:1\(2\)](https://doi.org/10.1061/(ASCE)0899-1561(2008)20:1(2)).
  - [80] A. Dousti, R. Rashednia, B. Ahmadi, M. Shekarchi, Influence of exposure temperature on chloride diffusion in concretes incorporating silica fume or natural zeolite, *Constr. Build. Mater.* 49 (2013) 393–399, <https://doi.org/10.1016/j.conbuildmat.2013.08.086>.
  - [81] 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. Civ. Eng.* 25 (2010) 242–253, [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000088](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000088).
  - [82] 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, <https://doi.org/10.1016/j.cemconres.2018.09.006>.
  - [83] 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 Journal of Materials in Civil Engineering* 31 (2019) 04019255, [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0002902](https://doi.org/10.1061/(ASCE)MT.1943-5533.0002902).
  - [84] T. Han, A. Siddique, K. Khayat, J. Huang, A. Kumar, An ensemble machine learning approach for prediction and optimization of modulus of elasticity of recycled aggregate concrete, construction and building materials, *Article in Press* (2020) 1–18, <https://doi.org/10.1016/j.conbuildmat.2020.118271>.
  - [85] 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, <https://doi.org/10.1016/j.eswa.2014.08.048>.
  - [86] H. Ma, Z. Li, Multi-aggregate approach for modeling interfacial transition zone in concrete, *ACI Mater. J.* 111 (2014) 189–200, <https://doi.org/10.14359/51686501>.