Composite Structures 284 (2022) 115184

Contents lists available at ScienceDirect

Composite Structures

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

Long-term performance prediction framework based on XGBoost decision tree for pultruded FRP composites exposed to water, humidity and alkaline solution

Xing Liu [a], TianQiao Liu [b], Peng Feng [a,*]

[a] Department of Civil Engineering, Tsinghua University, Beijing, China
[b] Department of Civil Engineering, Beijing University of Technology, Beijing, China

ARTICLE INFO

Keywords:
FRP composite materials
Long-term environmental effects
Machine learning analysis framework
XGBoost decision tree
Predictive model
Mechanical properties

ABSTRACT

Fiber reinforced polymer (FRP) composites are susceptible to material degradation when exposed to environmental effects. To predict the residual tensile strength and modulus of pultruded FRP composites, an XGBoost decision tree model was developed in this work. XGBoost decision tree, as a machine learning technique, is able to provide accurate predictions for tabular dataset with a good prediction interpretability. In this work, the methodology of XGBoost decision tree was presented in detail. Datasets for training and testing included a total of 746 data points which were collected from an existing database. XGBoost decision tree model predictions were cross-validated with 149 test data, and an excellent agreement was observed, showing R^2 values of 0.93 and 0.85 for tensile strength and modulus, respectively. In addition, attribute importance analysis was conducted to quantitatively evaluate the attributes pertaining to FRP degradations, including exposure time, exposure temperature, pH value of environment, fiber volume fraction, plate thickness, fiber type and matrix type. Exposure time and temperature were observed to have the greatest impacts on residual tensile properties. The proposed XGBoost decision tree model provides a new approach for predicting the long-term degradations of FRP composites subjected to environmental effects.

1. Introduction

Pultruded fiber reinforced polymer (FRP) composite materials have been increasingly used in the field of civil engineering [40,26,74,49]. Pultruded FRPs are well known to have high strength- and stiffness-to-weight ratios, making them a competitive alternative to conventional materials such as concrete and steel. In addition, FRP composites generally have better corrosion resistance over other materials such as steel, thus permitting the replacement of steel in those structures servicing in harsh environments, such as the chemical and marine environments. Nonetheless, FRP composites are not completely immune from material degradation when exposed to long-term environmental conditions. For instance, marine structures are highly susceptible to seawater ingression, and highway bridges are often exposed to high humidity, deicing salt and possibly acidic solutions [70]. All these environmental effects can negatively impact the long-term performance of FRP structures. The present authors [50] conducted a comprehensive review on eight environmental effects, including water/moisture, alkaline solutions, acidic solutions, low/high temperatures, ultraviolet radiation, freeze–thaw cycles, wet-dry cycles, and in-situ environments, and their degradation mechanisms may include both the physical and chemical reactions such as plasticization and hydrolysis. Given that FRP composites are often used in harsh environments, the environmental durability and the corresponding design method must be addressed to ensure the safe use of FRP structures.

Many pioneering studies have been carried out with the purpose of developing predictive models for FRP durability, and those benchmark studies are reviewed in this section. Arrhenius [5] established a phenomenal relationship between the temperature and the rate of chemical reactions in the late 1800s, and this relationship was widely adopted to simulate the material degradation of FRP composites subjected to a single environmental condition [18,27,51]. In addition to the Arrhenius model for a single environmental effect, Park et al. [64], based on the cumulative damage method and the stochastic process, proposed a hyper-cuboidal volume model to characterize material degradation due to the synergistic effect of multiple environmental

* Corresponding author.
E-mail addresses: liuxing17@mails.tsinghua.edu.cn (X. Liu), liu_tianqiao@bjut.edu.cn (T. Liu), fengpeng@tsinghua.edu.cn (P. Feng).

https://doi.org/10.1016/j.compstruct.2022.115184
Received 9 July 2021; Received in revised form 26 December 2021; Accepted 2 January 2022
Available online 6 January 2022
0263-8223/

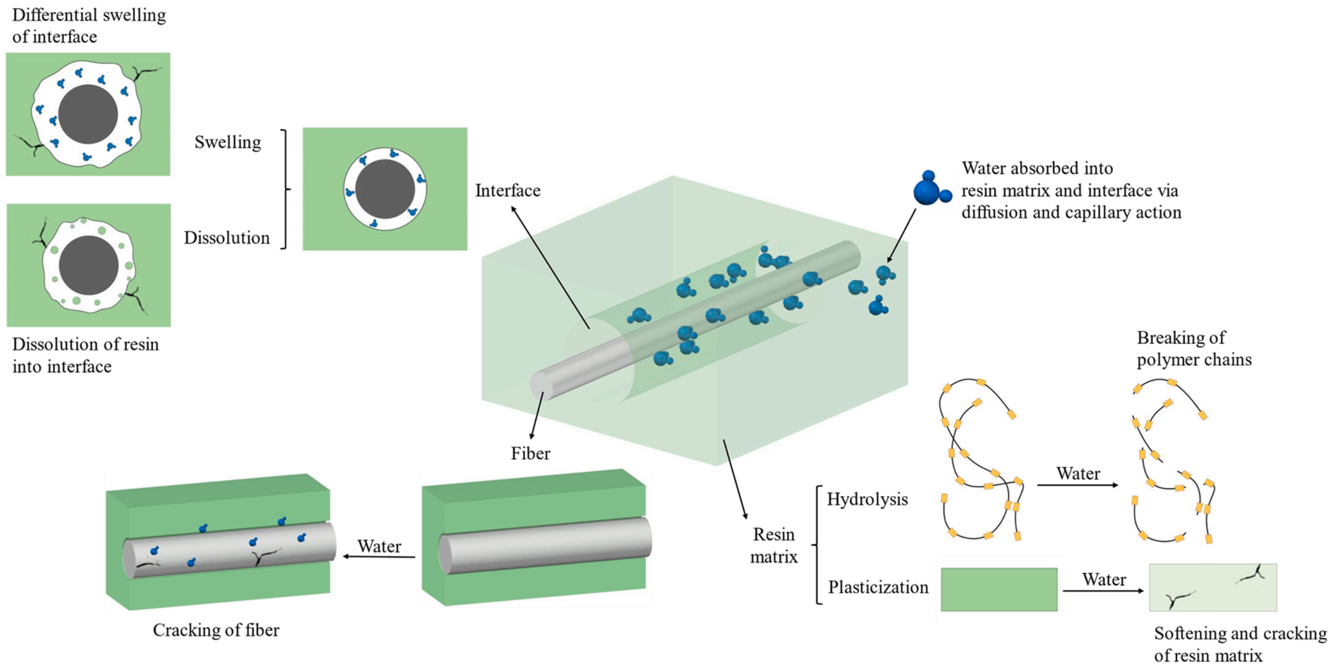

Fig. 1. Degradation mechanisms of FRP composites subjected to water/moisture ingression [50]: Water ingression could induce damage at all three phases of FRPs, including the fiber, matrix and fiber–matrix interface. First, microcracks can occur at the surface of the fibers and eventually lead to fiber rupture. Second, water ingression, realized through diffusion and capillary action, could cause swelling of the matrix and propagation of microcracks, thereby affecting the microstructure of the resin matrix. Typically, physical and chemical damage—plasticization and hydrolysis—can occur. Third, the water absorbed along the fiber–matrix interface could cause differential swelling of the interface and propagation of microcracks. Fiber-matrix debonding can occur, and a portion of matrix material may hydrolytically dissolve into the water.

conditions. In their model, the synergistic effect was assumed to be a multiplicative effect of each condition. Last but not least, the moisture/water ingression has been identified as one of the most detrimental effects on FRP composites [50,53,8], and accordingly, Fick's law draws great attention and has been widely adopted to define the moisture absorption of FRP composites. Based on Fick's law, many models have been developed to calculate the moisture content and diffusion coefficient of FRP composites [81,9,53]. However, it must be noted that both the Arrhenius relation and Fick's law were not originally derived for FRP degradation; that is, the use of such models essentially stems from the prediction ability of their mathematical expressions, rather than their mechanical, physical, or chemical explanation to FRP degradation. Additionally, the accuracy of these models is highly affected by the determination of calibration factors with regard to the experimental tests, and these models require a continuing correction or modification as new test data becomes available. This type of issue is also noted by Jiang et al. [41] who developed an artificial neural network model for predicting the stress–strain relation of FRP-confined concrete. Due to the inadequacy of predictive models, all the commonly accepted design guides [6,7,21] are still using empirically determined reduction factors, though these factors may differ from one guide to another and may greatly under-estimate the true durability of FRP composites.

The difficulty in deriving the universally reliable design equations for FRP durability is mainly due to three reasons [50]. First, the long-term performance of FRP composites is affected by a number of external environmental effects as well as by different compositions of materials, such as FRPs made of different fibers and resins with different fiber volume/mass ratios [2,58]. For instance, the degradation mechanism of FRPs subjected to water/moisture ingression is illustrated in Fig. 1. It is seen that water ingression may lead to damages on three phases of the material, including fiber, matrix and fiber–matrix interface, and the damages may be due to physical and chemical reactions, such as plasticization and hydrolysis [62,50,46]. These damages may occur at varying levels depending on the different types of fiber and matrix materials, the fiber volume fractions and the plate thickness of the FRP material being investigated. Such a complexity of FRP degradation is the first and foremost reason to hinder the development of design equations. Second, it is the synergistic effect of multiple environmental conditions that actually affects the long-term performance of FRP composites in the real world, and the synergistic effect greatly increases the complexity of the degradation mechanism of FRPs. For instance, the hygrothermal effect is a combined effect of high humidity and high temperature, and the resulting damage is not a simple linear accumulation of each single effect. Moreover, marine environment, as a typical harsh environment in civil engineering, consists of four different effects, including alkaline water ingression, ultraviolet radiation, wet-dry cycle and high temperature [77,78,1]. The resulting synergistic effect cannot be accurately assessed by existing predictive models. Finally, many analytical studies on durability of FRP composites are conducted at the microscopic molecular level, and the obtained findings are intrinsically difficult to be expanded to the macroscopic structural level [72]. The discrepancy between molecular and structural levels may result from the presumed simplifications in the molecular model and many other assumptions used for filling in the gap between microscopic and macroscopic levels. These three reasons, together, make the prediction of the lifetime mechanical properties of FRP composites an extremely challenging task.

The degradation of FRP composites is shown to be a highly nonlinear multi-dimensional problem with many influential factors. These influential factors may be categorized into the internal factors, such as the different fiber and matrix types and different fiber volume fractions, and the external factors, including different types of environmental conditions. In this regard, the conventional mechanical analysis can only help qualitatively understanding the long-term performance of FRP composites, whereas the accurate predictions of long-term mechanical properties of FRP composites can hardly be obtained. Thus, there is a gap between the understanding and the application of FRP composites. In order to fill this gap, a framework of long-term performance

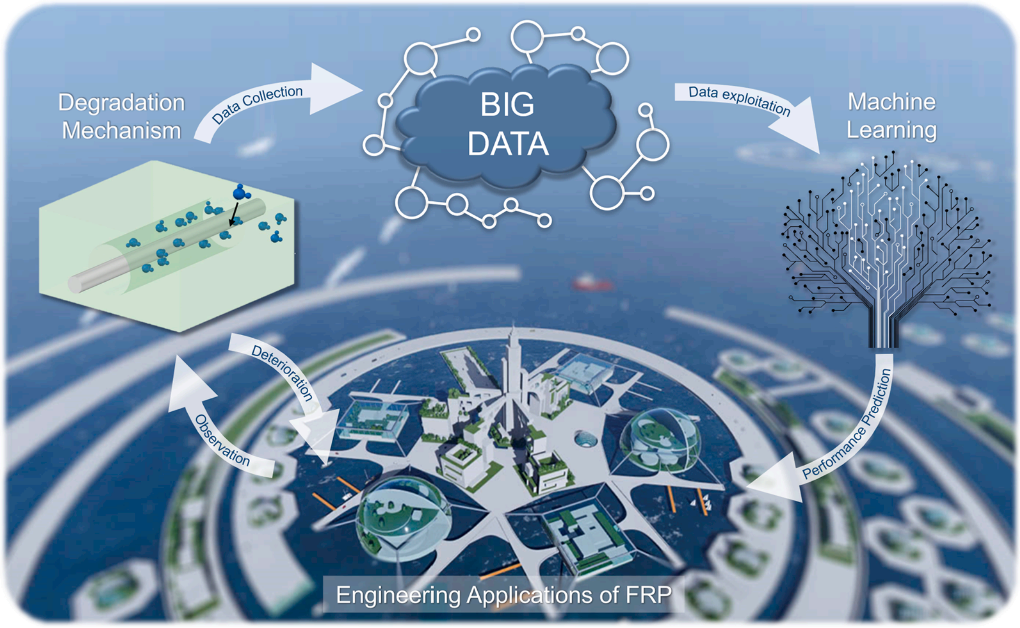

Fig. 2. Framework of long-term performance prediction of FRP composites: FRP deterioration is observed through numerous engineering applications (marine environment, for instance); then, the material degradation mechanism is studied, providing the big data in terms of the residual mechanical properties of FRPs; in this study, the big data is input into the machine learning model, which can provide accurate predictions of long-term performance of FRPs.

prediction of FRP composites is proposed in this work, as shown in Fig. 2. The big data on FRP degradation (seawater ingression, for instance) is first collected through numerous studies and applications and then, it is input into the machine learning model, which can provide accurate predictions of mechanical properties of FRPs.

Machine learning technique has tremendously progressed in the recent decades, and its applications have expanded beyond the scope of computer science, as it can provide a novel approach in solving conventional engineering problems. For instance, machine learning, combined with big data, has seen some pioneering applications in the field of civil engineering, such as predicting the performance of materials and monitoring the health of structures [71]. In particular, machine learning has been demonstrated to have many significant advantages in solving nonlinear regression and classification problems, particularly when a number of parameters are of interest. Additionally, machine learning can extract the internal correlation between input data and output results through continuous training, and its prediction ability can be further improved by enriched database and cross-validations. Machine learning, as the main category of its type, composes many different algorithms, and in this work, the decision tree model is selected to develop the predictive model for mechanical properties of pultruded FRP composites.

In this work, the longitudinal tensile properties of pultruded FRPs are focused as they are of the highest interest for many practical applications, and the corresponding feasibility study can readily serve as a guide for future studies for other mechanical properties, such as the compressive and shear properties. Moreover, the water, high humidity and alkaline solution are often considered the most detrimental conditions for FRP composites, and thus, they are selected as the objective environmental effects in this work. The database developed by Liu et al. [50] includes over 1900 data points directly obtained from material aging tests, and over 700 data points pertaining to tensile properties are selected for the purpose of this work. In the following sections, the decision tree model is first introduced and discussed in detail. Then, the predictive model is developed, and the predictions are validated with experimental results. The proposed model provides a novel approach for predicting the mechanical properties of FRP composites subjected to long-term environmental effects.

2. Decision tree model for civil engineering

To date, many machine learning algorithms have been adopted to solve problems in civil engineering [73]. The most commonly used algorithms may include artificial neural networks (ANNs), support vector machines (SVMs) and decision trees. Some studies [65,66,15,67] built ANN models to predict the shear and compressive strength of FRP bar-reinforced concrete beams and FRP-confined concrete columns. On the other hand, some studies [61,35,20] developed SVM and its derivative models to predict the compressive strength of normal and high-performance concrete, and FRP-confined concrete. Based on these studies, it is found that the ANN algorithm typically requires a relatively large database and a high computational cost for the training process so as to assure a satisfactory accuracy, and the SVM model with hyper-parameters can be properly trained only when the kernel functions, regularization penalties, and slack variables are all correctly selected. Thus, the applicability of ANN and SVM algorithms may be limited. The decision tree algorithm, compared to ANN and SVM, can provide a more robust approach. For instance, Marks et al. [56] built a decision tree to assess the surface scaling resistance of fly ash concrete and Mansouri et al. [54] proposed an improved M5 decision tree to predict the strength and FRP-confined concrete. Their study revealed the good result interpretability and accuracy of the decision tree model as compared to the ANN model. Moreover, Chou et al. [23] combined the regression tree and multilayer perceptron neural network via ensemble learning method to predict the compressive strength of high-performance concrete and found that the ensemble learning technique outperformed any single learning technique. In this regard, the boosting tree model, as a type of ensemble decision tree model, has drawn attention in the field as it can generate precise predictions by integrating the outputs from many weak tree models [82]. In fact, a single decision tree model may not necessarily outperform conventional neural network models, whereas the boosting tree model is able to generate much more accurate results than the ANN models [79,22]. The typical mathematical expression of the boosting tree model is given as [29]:

$$f_M(\mathbf{x}) = w_0 + \sum_{m=1}^{M} w_m \phi_m(\mathbf{x}) \tag{1}$$

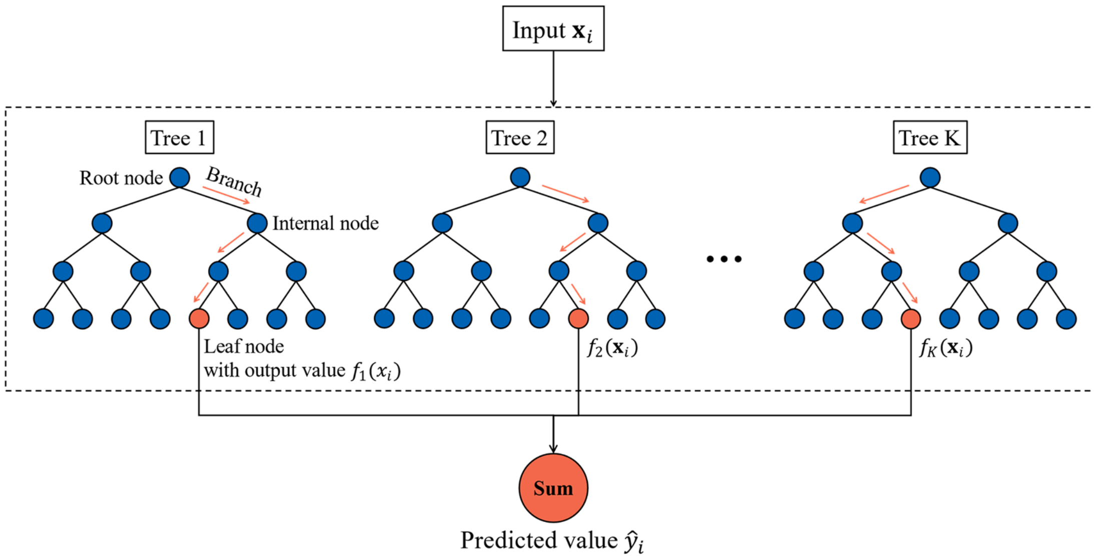

Fig. 3. Schematic of XGBoost decision tree model: Input data is first sent to the root node for initial decision; then, the internal nodes are to make the following decisions, and the branches direct to the decisions to be processed; the leaf node yields the prediction of a tree; finally, the predictions from all trees are added up, and the final prediction is obtained.

where $\mathbf{x}$ is the input variable; M is the total number of decision trees in the boosting model, and each tree is a weak prediction model; and $\phi_m(\mathbf{x})$ and w_m are the prediction result and the weight of the mth tree, respectively. Thus, the overall prediction result of boosting tree $f_M(\mathbf{x})$ is the weighted sum of all decision trees.

Among all types of boosting trees, gradient boosting decision tree (GBDT) is often deemed the most representative boosting tree algorithm. GBDT uses gradient descent to minimize the loss function in each step and generate a new decision tree. Chou et al. [22] compared the prediction accuracy of concrete compressive strength predicted by five different algorithms, including the ANN, SVM, multiple regression, bagging regression tree, and GBDT. GBDT was seen to have powerful potential in achieving the best accuracy. In addition, conventional GBDT can be further upgraded to an extreme gradient boosting (XGBoost) decision tree. In recent years, XGBoost has shown outstanding performance in numerous data mining competitions, and its prediction output can be readily interpreted. The advantages of XGBoost stem from its new features, such as structure penalization of trees, random variables and parallel calculation abilities [17]. The representative applications of XGBoost decision tree can be found in Dong et al. [30] and Lim and Chi [48] on material and structural levels, respectively. Recently, Duan et al. [31] compared the performance of XGBoost, ANN, and SVM in predicting the compressive strength of recycled aggregate concrete, and XGBoost was reported to outperform the other algorithms. Based on findings from previous studies, it can be concluded that XGBoost decision tree has an excellent capability of solving the nonlinear regression problem.

The degradation of FRP composites subjected to long-term environmental effects is highly nonlinear, and in this work, XGBoost decision tree algorithm is adopted to predict the residual longitudinal tensile properties of pultruded FRPs. The proposed XGBoost model employs the GridSearchCV and k-fold cross-validation methods to determine the best hyperparameters of the algorithm. The statistical scores, including the R-square value (R^2), root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE), are taken as the criteria to evaluate the accuracy of predictions. Finally, the respective contribution/importance of each input variable (i.e., parameters of interest) is analyzed and visualized via attribute importance analysis.

3. Methodology

3.1. XGBoost algorithm

Considering the complex degradation mechanisms of pultruded FRP composites under long-term environmental effects, XGBoost algorithm was selected in this work to develop the predictive model for the mechanical properties of FRP materials after the comparison and trials of different algorithms. XGBoost, first proposed by Chen and Guestrin [17], is capable of analyzing the specific importance of a variety of parameters in the model. In addition to the well-acknowledged accuracy, XGBoost has many advantageous features. First, the input dataset is organized in a tabular form, which is indeed the preferred data form in civil engineering. Some studies [16,75,80] have clearly reported that XGBoost is more proficient in using tabular datasets than other algorithms, such as the ANN model that typically requires a large-scale dataset in forms of pictures and/or videos. Second, XGBoost is a sparsity-aware algorithm. This is of particular significance when some parts of data are missing. In this case, XGBoost can automatically find the optimal results for the missing data based on the rest of the dataset. Third, XGBoost prescribes a high penalty on the structural complexity of the model, and this mechanism can efficiently prevent overfitting. Finally, XGBoost works faster with regard to the training process than other tree models since it can perfectly organize the software platform and hardware resources and generate each tree instance by parallel computing.

The XGBoost algorithm is composed of many classifications and regression trees (CARTs) that are capable of solving both classification and regression problems. In this work, the residual longitudinal tensile properties of pultruded FRP composites are to be predicted, which can be considered a regression problem. The structure of XGBoost can be expressed in the form of a flowchart, including multiple root nodes, many internal nodes, branches, and leaf nodes, as shown in Fig. 3. In this structure, the ith parameter $\mathbf{x}_i$ is input into the model and delivered to all the root nodes of all CARTs for the initial decisions. Then, the internal nodes make the following decisions. The branches direct to the decisions to be processed. The leaf nodes represent the prediction results of each single CART. Finally, all the results from leaf nodes are combined together, yielding the prediction result of XGBoost model [13]. Taking the ith dataset ($\mathbf{x}_i$, y_i) as an example ($\mathbf{x}_i$ is the input variable with several attributes and y_i is the real value for validation purpose), XGBoost model can be mathematically expressed using Eq. (2) [17].

$$\widehat{y}_i = \sum_{m=1}^{M} f_m(\mathbf{x_i}) \tag{2}$$

where $\widehat{y}_i$ is the predicted value with respect to input $\mathbf{x_i}$; M is the total number of CARTs being used; and f_m represents the predicted value of each independent CART. Eq. (2) clearly shows that the predicted score $\widehat{y}_i$ with respect to input $\mathbf{x_i}$ is given as the sum of all f_m values.

With the prediction result being obtained, an objective function is needed to evaluate the quality of result. In XGBoost algorithm, the objective function L is given as [63]:

$$L = \sum_{i}^{n} l(y_i, \widehat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{3}$$

This objective function contains two parts: 1) the loss function l, which measures the distance between predicted value $\widehat{y}_i$ and real value y_i; and 2) the regularization item Ω, which penalizes the complexity of tree structure. The specific form of Ω for one CART is given as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{4}$$

where T is the total number of leaf nodes of a CART; w_j is the predicted value of the jth leaf node; and γ and λ are the hyperparameters of the algorithm. When increasing γ and λ, the penalty for structural complexity of CART is increased; that is, increasing the complexity of tree leads to increased penalty. The goal of Ω is to make XGBoost a simple tree structure as well as to avoid overfitting.

In order to minimize the objective function and achieve the best prediction result, training of XGBoost model is needed. This process, often known as the optimization process, is conducted in a step-by-step manner. In each step, a new CART is generated based on existing CARTs, and the objective function is further reduced. The objective function of the tth step $L^{(t)}$ can be calculated based on the previous step $L^{(t-1)}$ as:

$$\begin{aligned} L^{(t)} &= \sum_{i}^{n} l\left(y_i, \widehat{y}_i^{(t)}\right) + \sum_{i=1}^{t} \Omega(f_i) \\ &= \sum_{i}^{n} l\left(y_i, \widehat{y}_i^{(t-1)} + f_t(x_i)\right) + \sum_{i=1}^{t-1} \Omega(f_i) + \Omega(f_t) \end{aligned} \tag{5}$$

The existing $(t-1)$ CARTs are known and can be seen as a constant in the tth step; that is, the second term on the right-hand side of Eq. (5) can be replaced by a constant c. Then, $L^{(t)}$ can be simplified as:

$$L^{(t)} = \sum_{i}^{n} l\left(y_i, \widehat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) + c \tag{6}$$

Additionally, by applying the second-order Taylor expansion to above equation, the objection function can be transformed into:

$$L^{(t)} = \sum_{i=1}^{n} \left[l\left(y_i, \widehat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) + c \tag{7}$$

where

$$g_i = \frac{\partial l\left(y_i, \widehat{y}_i^{(t-1)}\right)}{\partial \widehat{y}_i^{(t-1)}} \tag{8}$$

$$h_i = \frac{\partial^2 l\left(y_i, \widehat{y}_i^{(t-1)}\right)}{\partial \left(\widehat{y}_i^{(t-1)}\right)^2} \tag{9}$$

When optimizing the tth CART, there are n pairs of g_i and h_i to be calculated (n is the total number of datasets). Given that each input variable is independent, g_i and h_i can be calculated in a parallel manner. With that said, the CPU resources can be fully utilized, and the calculation speed can be greatly improved.

In addition, the form of loss function l can be determined according to specific problem. The only requirement is that the loss function must permit the second-order derivative. In this work, the expression of residual standard error (RSE) is selected as the loss function. Since each input variable $\mathbf{x_i}$ is to be projected to a leaf node of a CART, $f_k(\mathbf{x_i})$ can be written as:

$$f_k(\mathbf{x_i}) = \omega_{q(\mathbf{x_i})}, \omega \in R^T, q : R^d \rightarrow \{1, 2, \cdots, T\} \tag{10}$$

where $q(\mathbf{x_i})$ is a function that maps $\mathbf{x_i}$ (i.e., d-dimensional vector) to the index of a specific leaf node; w is the value of this specific leaf node; T is the leaf node number of the kth tree; d is the attribute number of the input $\mathbf{x_i}$; and R^T and R^d indicate T-dimensional and d-dimensional vectors. Substituting Eqs. (4), (8), (9), and (10) into Eq. (7) yields:

$$\begin{aligned} L^{(t)} &\approx \sum_{i=1}^{n} \left[g_i \omega_{q(\mathbf{x_i})} + \frac{1}{2} h_i \omega_{q(\mathbf{x_i})}^2 \right] + \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 + c \\ &= \sum_{j=1}^{T} \left[\left(\sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left(\sum_{i \in I_j} (h_i + \lambda) \omega_j^2 \right) \right] + \gamma T + c \end{aligned} \tag{11}$$

Letting $G_j = \sum_{i \in I_j} g_i$ and $H_j = \sum_{i \in I_j} h_i$, Eq. (11) is simplified as:

$$L^{(t)} = \sum_{j=1}^{T} \left[G_j \omega_j + \frac{1}{2} (H_j + \lambda) \omega_j^2 \right] + \gamma T + c \tag{12}$$

In order to determine the minimum value of objective function L, the first derivative of Eq. (12) is obtained. L_{min} is therefore calculated as:

$$L_{min} = \frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T + c \tag{13}$$

and L_{min} is obtained when ω_j is taken as:

$$\omega_j = -\frac{G_j}{H_j + \lambda} \tag{14}$$

The minimum value of objective function L is, therefore, the predicted value shown on leaf node. Furthermore, to find the best structure of each CART, a greedy algorithm is adopted to optimize the tree structure [33]. In particular, a gain function is used, as shown in Eq. (15).

$$Gain = \frac{1}{2} \left(\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right) - \gamma \tag{15}$$

The gain function has four terms: the first two terms are the profits of left and right parts of a node, where G_L, G_R are the left and right parts of G_j, and H_L and H_R are the left and right parts of H_j; and the third item is the total profit of that node. The last item is the regularization item for preventing overfitting. The greedy algorithm determines whether a node obtains the maximum gain. Thus far, the optimal tree structure that maximizes the gain can be generated.

3.2. Performance evaluations

In order to evaluate the performance of the proposed model, four statistical criteria, including R^2 value, RMSE, MAE and MAPE, are used. Their mathematical expressions are presented in Eqs. (16)–(19).

$$R^2 = 1 - \frac{\sum_i (\widehat{y}_i - y_i)^2}{\sum_i (y_i - \overline{y})^2} \tag{16}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_i (y_i - \widehat{y}_i)^2} \tag{17}$$

$$MAE = \frac{\sum_i |y_i - \widehat{y}_i|}{\sum_i |y_i - \overline{y}|} \tag{18}$$

$$MAPE = \sum_{i} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \tag{19}$$

where y_i is the real value; $\hat{y}_i$ is the predicted value; and $\bar{y}$ is the average of real values. R^2 measures the correlation between real value and predicted value, and a greater R^2 indicates a better performance of the model. The other three methods (i.e., RMSE, MAE and MAPE) measure the distance between real value and predicted value, and smaller RMSE, MAE and MAPE indicate a better performance.

3.3. Prediction interpretations

A good machine learning algorithm can not only provide good prediction accuracy, but also have good interpretability on the predicted results [43,59]. In general, an algorithm with good interpretability can facilitate the comprehension and acceptance of the model being developed. In this work, attribute importance analysis is conducted to assess the interpretability of proposed model. Attribute importance is used to directly quantify the specific contribution/importance of each attribute of a variable in dataset (considering one variable may have multiple attributes). In XGBoost, the attribute importance of a variable can be explicitly expressed as a score. Those attributes with high scores are expected to have a greater impact on the final predictions.

4. Analytical program

In this work, an XGBoost decision tree-based predictive model is developed to calculate the residual tensile strength and modulus of pultruded FRP composites subjected to water immersion and alkaline solution immersion. This model enables the analysis of all the possible influential parameters pertaining to pultruded FRP composites and environmental conditions. The detailed developments of proposed model are presented in this section.

4.1. Dataset determination

Machine learning, as a data analysis technique, is carried out on an existing database. The review work conducted by the present authors [50] summarized over 1,900 experimentally obtained residual properties of FRP composites subjected to eight environmental conditions, including water/high humidity, alkaline solutions, acidic solutions, low/high temperature, ultraviolet radiation, freeze–thaw cycle, wet-dry cycle, and in-situ environment. The XGBoost model addressed in this work is constructed based on this database. In particular, this work is focused on the tensile properties of FRPs exposed to water, high humidity, and alkaline solutions as these conditions have the most detrimental effects on FRPs [28,55]. Moreover, focusing on alkaline solutions also stems from the fact that pultruded FRP composites are being increasingly used in marine structures and alkaline environment is of critical design concern [70]. In addition, four rules are adopted for selecting the appropriate data for the XGBoost decision tree model:

(1) Tests with quantified results are selected, while tests that do not include quantified results are filtered out.
(2) Tests conducted on FRP composites alone are selected, while tests carried out on FRP-concrete hybrid materials are filtered out, such as concrete beams reinforced with FRP bars.
(3) Tests conducted on thermoset resin-based materials are selected, including epoxy-, polyester-, and vinyl ester-based materials, while tests on thermoplastic resin-based materials are filtered out.
(4) Tests conducted on carbon, glass and basalt fiber-reinforced materials are selected, while tests on organic fiber-reinforced materials, such as aramid FRPs, are filtered out.

With all the data being filtered, a total of 275 data points (i.e., tested mechanical properties) under water immersion/high humidity conditions and 267 data points under alkaline solution conditions are obtained to develop the predictive model for the tensile strength of pultruded FRP composites. On the other hand, a total of 118 data points under water/high humidity conditions and 86 data points under alkaline solution conditions are extracted to construct the predictive model for the tensile modulus of elasticity of pultruded FRP composites. It is noted that the repeated test results from different studies, such as Chen et al. [18] and Chen et al. [19], and Chu et al. [24] and Chu and Karbhari [25], are recognized and only counted once.

These data are deemed the raw dataset and contain all necessary information regarding materials and tests, including the composition of materials (i.e., fiber type, resin type and amount of fiber/resin content) and the specific environmental conditions (i.e., exposure time and temperature, and pH value of alkaline solution). In order to construct the decision tree, all of the information needs to be parameterized. The rules of parameterization are as follows:

(1) Different fibers and resins have different mechanical properties, and thus, they are considered different types of materials in this model. To differentiate and parameterize the fibers and resins, 3-dimensional vectors are used. Following the method by [37,38], FRPs with glass (G), carbon (C), and basalt (B) fibers are parameterized as a vector of (v_G, v_C, v_B). Similarly, FRPs with epoxy (E), polyester (P), and vinyl ester (V) resins are parameterized as a vector of (v_E, v_P, v_V). In addition, 1 and 0 are used to denote the status of fiber and resin. For instance, an FRP made of glass fiber and polyester resin is denoted as (1, 0, 0) for fiber and (0, 1, 0) for resin.
(2) Different pultruded FRP profiles exist in the database, including FRP plates (whose size is described by plate thickness) and FRP bars (whose size is described by bar diameter). To unify the thickness of all FRP materials, the diameter of bars is taken as the thickness of materials.
(3) The pH value of fresh/distilled/tap/deionized/demineralized water is taken as 7, and the pH values of alkaline solutions are taken as those prescribed in the tests. Provided that the specific pH value is absent in the test, strong alkaline solutions are set to pH = 13, and weak alkaline solutions are set to pH = 8. In addition, the on-site/artificial seawater environment is set to pH = 8.
(4) To unify and parameterize the durations of accelerating tests, the test time is converted to hours. The test time of all control specimens is naturally denoted as zero.
(5) The temperatures of all accelerating tests are converted to Celsius. In addition, room temperature is considered to be 20 ℃.
(6) The relative humidity is considered to be 0.46 in absence of information. For water condition, the humidity is considered to be 1.

Through data preprocessing, all the information of data are parameterized into a structured vector. Since the XGBoost model can automatically fill in for those missing items in the vector, the absent data can be left blank and set to null when training the model [52].

4.2. Model development

The preprocessed data are then randomly grouped into 1) a training set and 2) a testing set at an empirical ratio of 4:1 [11]. Random grouping can remove the interferences from all other possible external factors of the dataset. In addition, in this work, the water and high humidity conditions are grouped as they have similar degradation effect on FRPs; that is, those data points pertaining to high humidity are included in the dataset of water condition. The pH value is used as a criterion for differentiating between water and alkaline solutions in the decision tree,

Table 1
Information availabilities of collected dataset.

Model type	Environmental condition	Author	Test information availabilities								Number of collected data points
			Fiber type	Matrix type	Fiber volume fraction	Plate thickness	pH value	Relative humidity	Exposure time	Exposure temperature	
S-model	Alkaline solution immersion	Gentry et al. [34]	√	√	√	√	√	–	√	√	11
		McBagonluri et al. [57]	√	√	√	√	√	–	√	√	1
		Chu et al. [24]	√	√	√	√	√	–	√	√	29
		Micelli and Nanni [58]	√	√	√	√	√	–	√	√	9
		Chen et al. [19]	√	√	√	√	√	–	√	√	16
		Kafodya et al. [42][1]	√	√	–	√	–	–	√	√	12
		Lu et al. [51]	√	√	√	√	√	–	√	√	36
		Heshmati et al. [39]	√	√	√	√	–	–	√	√	8
		Chen et al. [18]	√	√	√	√	√	–	√	√	17
		Kim et al. [44]	√	√	√	√	√	–	√	√	65
		Won et al. [76]	√	√	–	√	√	–	√	√	25
		Sawpan et al. [69]	√	√	–	√	√	–	√	√	8
		Cabral-Fonseca et al. [14]	√	√	√	√	–	–	√	√	30
	Water immersion	Liao et al. [47]	√	√	√	√	√	–	√	√	2
		Shao and Kouadio [70]	√	√	√	√	√	–	√	√	20
		Gentry et al. [34]	√	√	√	√	√	–	√	√	12
		McBagonluri et al. [57]	√	√	√	√	√	–	√	√	2
		Chen et al. [19]	√	√	√	√	√	–	√	√	4
		Chu and Karbhari [25]	√	√	√	√	√	–	√	√	57
		Kafodya et al. [42]	√	√	–	√	√	–	√	√	13
		Lu et al. [51][2]	√	√	√	√	√	–	√	√	39
		Grammatikos et al. [36]	√	√	√	√	√	–	√	√	17
		Heshmati et al. [39]	√	√	√	√	√	√	√	√	14
		Kim et al. [44]	√	√	√	√	√	–	√	√	45
		Al-Salloum et al. [3]	√	√	√	√	√	–	√	√	7
		Cabral-Fonseca et al. [14]	√	√	√	√	√	–	√	√	38
		Zhang et al. [83]	√	√	√	√	√	–	√	√	5
M–model	Alkaline solution immersion	McBagonluri et al. [57]	√	√	√	√	√	–	√	√	1
		Micelli and Nanni [58]	√	√	–	√	√	–	√	√	9
		Kafodya et al. [42]	√	√	–	√	–	–	√	√	12
		Lu et al. [51][2]	√	√	√	√	√	–	√	√	9
		Heshmati et al. [39]	√	√	√	√	–	–	√	√	8
		Kim et al. [44]	√	√	√	√	√	–	√	√	39
		Sawpan et al. [69]	√	√	–	√	√	–	√	√	8
	Water immersion	Liao et al. [47]	√	√	√	√	√	–	√	√	2
		Shao and Kouadio [70]	√	√	√	√	√	–	√	√	20
		McBagonluri et al. [57]	√	√	√	√	√	–	√	√	2
		Kafodya et al. [42][1]	√	√	–	√	√	–	√	√	13
		Lu et al. [51][2]	√	√	√	√	√	–	√	√	12
		Grammatikos et al. [36]	√	√	√	√	√	–	√	√	17
		Heshmati et al. [39]	√	√	√	√	√	√	√	√	14
		Kim et al. [44]	√	√	√	√	√	–	√	√	26
		Al-Salloum et al. [3]	√	√	√	√	√	–	√	√	7

(continued on next page)

Table 1 (*continued*)

Model type	Environmental condition	Author	Test information availabilities								Number of collected data points
			Fiber type	Matrix type	Fiber volume fraction	Plate thickness	pH value	Relative humidity	Exposure time	Exposure temperature	
		Zhang et al. [83]	√	√	√	√	√	–	√	√	5

[1] Test specimens were also subjected to external loading.
[2] Test specimens were preheated before immersion.

Table 2
Attributes and data ranges of collected dataset.

Model type	Attributes of dataset [data range]						
	Fiber type	Matrix type	V_f	Coupon thickness (mm)	pH value	Exposure time (hours)	Exposure temperature (°C)
S-model	[G, C, B]	[E, P, V]	[0.22, 0.83]	[0.7, 14]	[7, 13]	[0, 20160]	[20, 100]
M−model	[G, C, B]	[E, P, V]	[0.29, 0.83]	[1.25, 14]	[7, 13]	[0, 20160]	[20, 100]

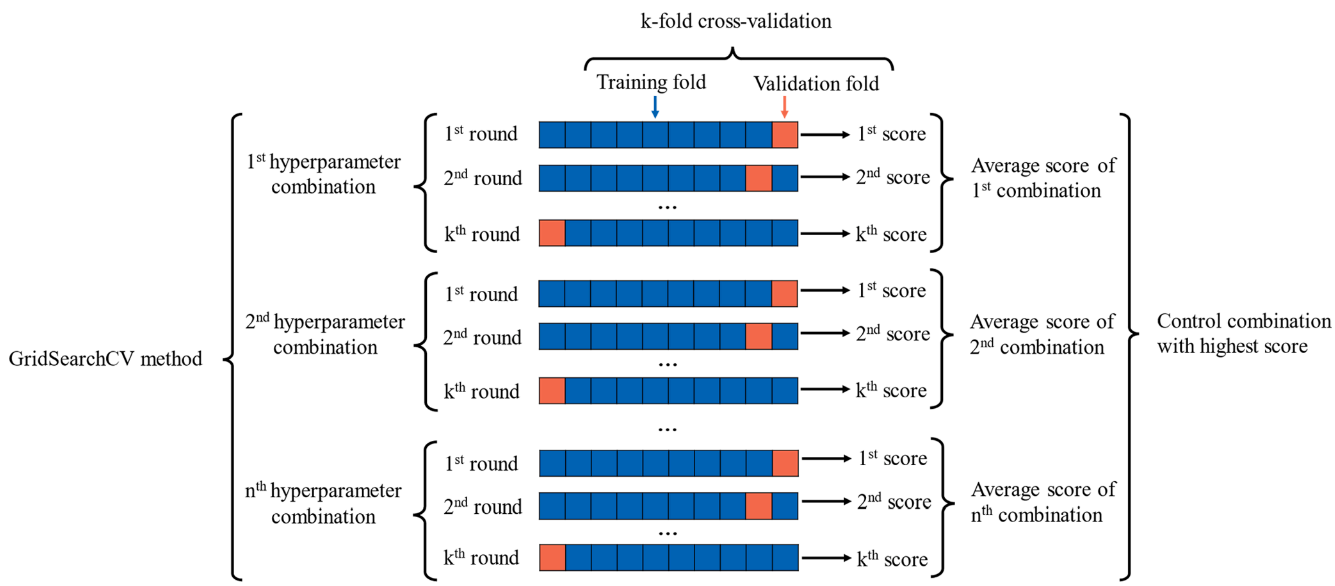

Fig. 4. Schematic of GridSearchCV method and k-fold cross-validation for determining control hyperparameter combination: GridSearchCV automatically evaluates all the possible combinations of hyperparameters, and the combination having the highest performance score is taken as the control solution. The performance score is calculated through k-fold cross-validation, which automatically divides the training dataset into a training fold and a validation fold at a ratio of (k-1):1 (k = 10, in this case). Then, this training and cross-validation process is repeated k times; in each round, the validation fold is shifted to another group, and R^2 value is calculated as the performance score. As a result, for each possible hyperparameter combination, a total of k scores is obtained, and the arithmetic average of these k scores is taken as the performance score of that hyperparameter combination.

which allows the emergence of two datasets. Therefore, for predicting the tensile strength, 275 water condition data points and 267 alkaline solution condition data points can be emerged, resulting in a total of 542 data points. This dataset for tensile strength is further divided into a training set (with 434 data points) and a testing set (with 108 data points). Similarly, for predicting the tensile modulus of elasticity, 118 water condition data points and 86 alkaline solution condition data points are emerged, yielding a total dataset with 204 data points. This dataset for tensile modulus is then divided into the training set (with 163 data points) and testing set (with 41 data points). The predictive models for tensile strength and modulus are designated as the S-model and M−model, respectively. The information availabilities of the collected dataset are presented in Table 1, and the data range of each attribute is showed in Table 2. As aforementioned, XGBoost can automatically fill in for missing information based on the remaining dataset.

With the training and testing datasets determined, the hyperparameters of XGBoost model are to be determined, including tree-number, learning-rate, max-depth, min-child-weight, subsample, colsample-bytree, γ, α, and λ. Tree-number indicates the number of regression trees in XGBoost model; learning-rate defines the step size of each training round; max-depth defines the number of branches from root node to leaf node of a regression tree, namely, the depth of a regression tree; min-child-weight defines the complexity of a regression tree (in this case, a smaller min-child-weight value results in a more complex tree model that is more likely to be overfitted); subsample indicates the ratio of training set to the total dataset of a regression tree; colsample-bytree indicates the ratio of training attributes to total attributes of a regression tree; and γ, α, and λ are regularization factors of the objective function (see Eq. (4)), which are used to prevent the model from overfitting.

Appropriate selections of hyperparameters directly impact the overall performance of XGBoost model. In order to find the optimal combination of hyperparameters, GridsearchCV method and k-fold cross-validation were adopted. This process is schematically shown in Fig. 4. GridSearchCV is an automated parameter tuning method, which is essentially an exhaustive search method; that is, all possible

Table 3
Hyperparameters of XGBoost decision tree model.

	Hyperparameters	Initial value	[Test range] (Increment size)	Control value	
				S-model	M−model
1.	Tree-number	50	[20, 500] (10)	400	150
2.	Learning-rate	0.1	[0.01, 0.05, 0.07, 0.1, 0.2, 0.5, 1, 2, 4]	0.1	0.1
3.	Max-depth	5	[3, 10] (1)	5	5
4.	Min-child-weight	6	[1, 10] (1)	4	5
5.	Subsample	0.8	[0.1, 1.0] (0.1)	0.7	0.9
6.	Colsample-bytree	0.8	[0.1, 1.0] (0.1)	0.8	0.6
7.	γ	0	[0, 0.6] (0.1)	0	0
8.	α	0	[0, 0.05, 0.1, 1, 2, 3, 4]	0	0
9.	λ	1	[0, 0.05, 0.1, 1, 2, 3, 4]	1	1

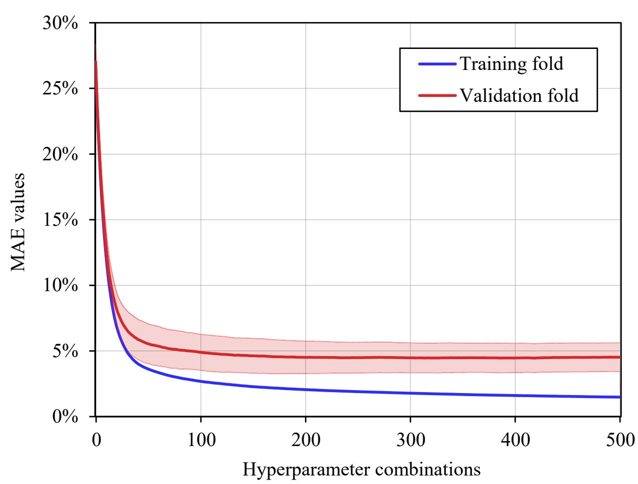

Fig. 5. MAE vs. hyperparameter combination (n = 500).

combinations of hyperparameters are to be evaluated and compared, and the combination having the highest performance score is taken as the control solution, namely, the control hyperparameter combination [12,68]. In this case, the total number of hyperparameter combinations is referred to as n. The test range of hyperparameters is shown in Table 3. Moreover, R^2 value is used as the performance score, as calculated through k-fold cross-validation. K-fold cross-validation automatically divides the training dataset into a training fold and a validation fold at a ratio of (k-1):1. Then, this training and cross-validation process is repeated k times; in each round, the validation fold is shifted to another group, and R^2 value is calculated as the performance score. As a result, for each possible hyperparameter combination, a total of k scores is obtained. The arithmetic average of these k scores is taken as the performance score of that hyperparameter combination; in this work, k is taken as 10 [45]. The combination with the highest score (among a total of n combinations) is therefore the control combination for XGBoost model.

In addition to R^2 value, the mean absolute error (MAE) is adopted as the second performance score so as to double-check the optimal hyperparameter combination. GridSearchCV method and k-fold cross-validation are, again, conducted, and in this process, the performance scores are all replaced by MAE values. A total of 500 hyperparameter combinations (i.e., $n = 500$ in this case) are tested. The obtained average MAE values of S-model are plotted against their corresponding hyperparameter combinations, as shown by the solid curves in Fig. 5. The shaded area represents ±1 standard deviation based on the average MAE value, and this standard deviation was calculated based on the 10 MAE values of each combination ($k = 10$ in this case). The MAE values of both the training fold and validation fold decrease as the number of combinations increases. This indicates that the prediction performance increases with increasing hyperparameter combinations. Additionally, MAE values of the training fold are lower than those of the validation fold; that is, the accuracy of training fold is higher than that of the validation fold. This is because only the training fold is involved in the training process, and the resulting predictions naturally have a better correlation with the training fold. Fig. 5 also shows that MAE values of the training fold start to converge at approximately 100 combinations. The control combination obtained from previous process (i.e., R^2 value-based process) is 400. Based on the results from both R^2 and MAE analysis, the number of hyperparameter combinations is finally taken as 400 for S-model. When repeating this MAE-based analysis for M−model, the optimal number of combinations is also obtained, which is 150. On the other hand, the control combination obtained from R^2 value-based analysis is 80. Based on those results, the control number of hyperparameter combinations is finally taken as 150 for M−model.

The control hyperparameters are presented in Table 3. With all hyperparameters determined, the XGBoost model is finalized, and the predictive model is successfully obtained. The flowchart of the entire development of predictive model is presented in Fig. 6. Using this model, the residual tensile properties of pultruded FRP composites subjected to water immersion and alkaline solution immersion are calculated. The results are presented in the following sections.

5. Prediction results and analysis

5.1. Prediction results

With the optimal hyperparameters determined via GridSearchCV method, predictive models for the tensile properties of pultruded FRP composites under environmental effects of water, high humidity and

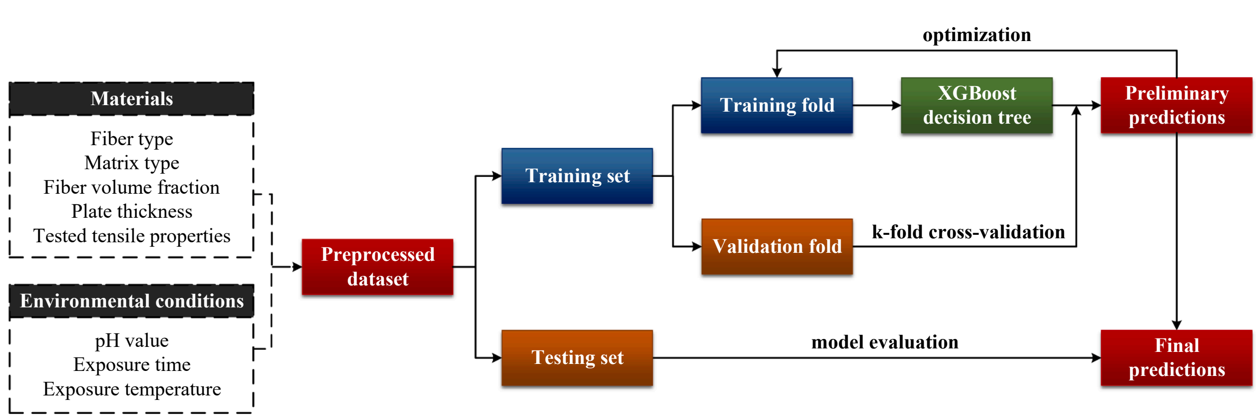

Fig. 6. Flowchart of predictive model development.

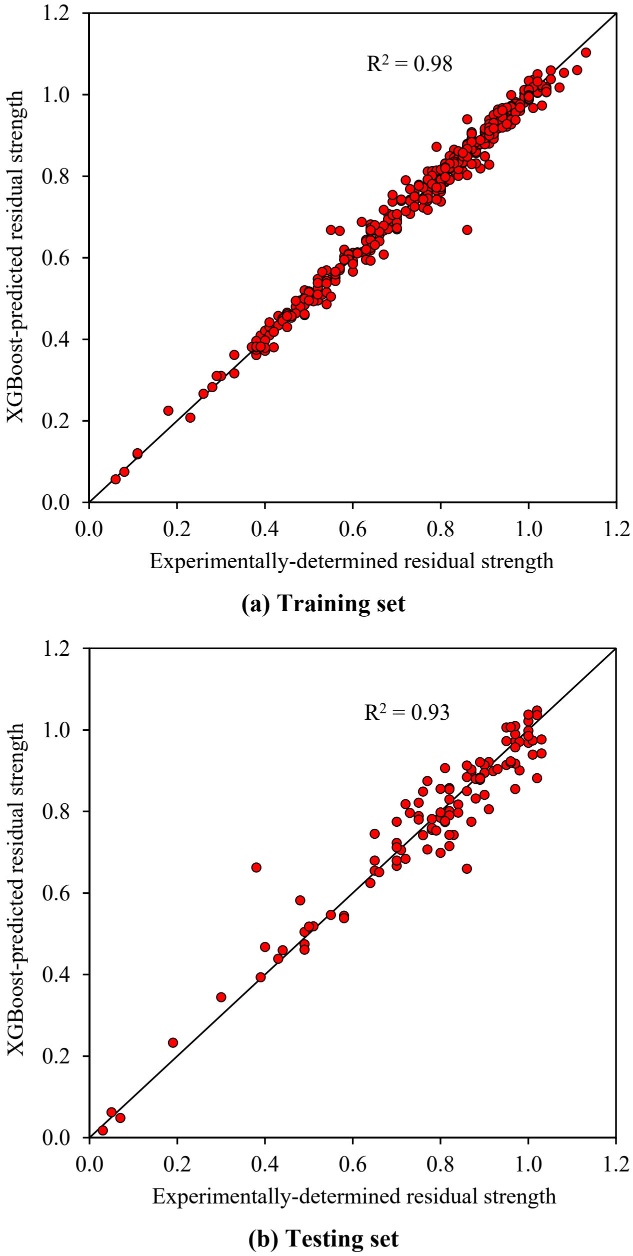

Fig. 7. XGBoost-predicted residual strength vs. experimentally-determined residual strength (strength normalized for simplicity).

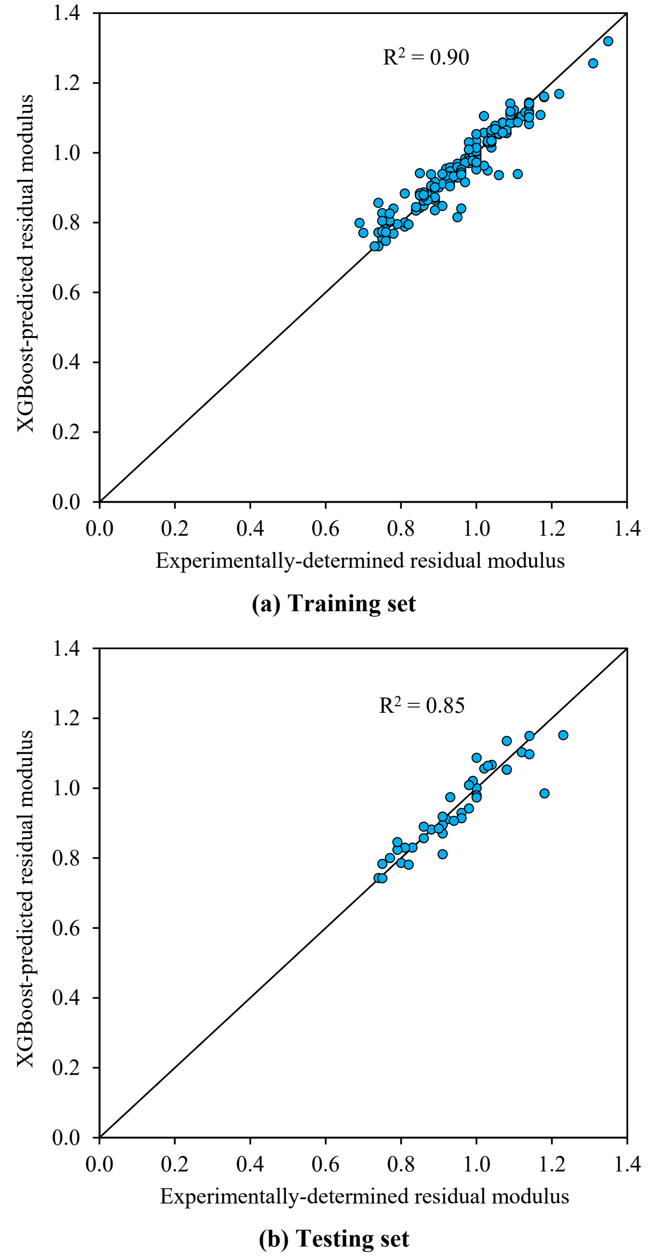

Fig. 8. XGBoost-predicted residual modulus vs. experimentally-determined residual modulus (modulus normalized for simplicity).

alkaline solution immersions are obtained, including S-model for tensile strength and M−model for tensile modulus of elasticity. The prediction results for testing set as well as the detailed information regarding the aging tests are presented in Appendix A. XGBoost-predicted residual tensile strength and experimentally determined residual tensile strength are all normalized against their as-received tensile strength, and the predicted strength is plotted against the experimental results, as shown in Fig. 7. For both the training set and testing set, XGBoost-predicted tensile strengths are generally close to those obtained from accelerated aging tests. The training set indeed shows a better correlation than testing set. This is actually expectable for most machine learning techniques since the training set is directly used to output the predictions. In addition, XGBoost-predicted residual tensile modulus of elasticity and experimentally determined residual tensile modulus are normalized against the as-received tensile modulus, and the predicted modulus are plotted against experimental results, as shown in Fig. 8. For training set, a good correlation between XGBoost predictions and experimental results can be seen, while for testing set, that correlation is reduced. This is mainly due to the smaller volume of dataset for M−model.

To quantify the performance as well as to evaluate the accuracy of XGBoost model, XGBoost predictions are evaluated via four evaluation criteria, including R^2 value, RMSE, MAE and MAPE. The evaluation results are calculated and presented in Table 4.

Table 4
Evaluation results of S-model and M−model.

Evaluation criteria	S-model		M−model	
	Training set	Testing set	Training set	Testing set
R^2	0.98	0.93	0.90	0.85
RMSE	0.02	0.06	0.04	0.05
MAE	0.02	0.04	0.03	0.03
MAPE	0.02	0.07	0.03	0.03

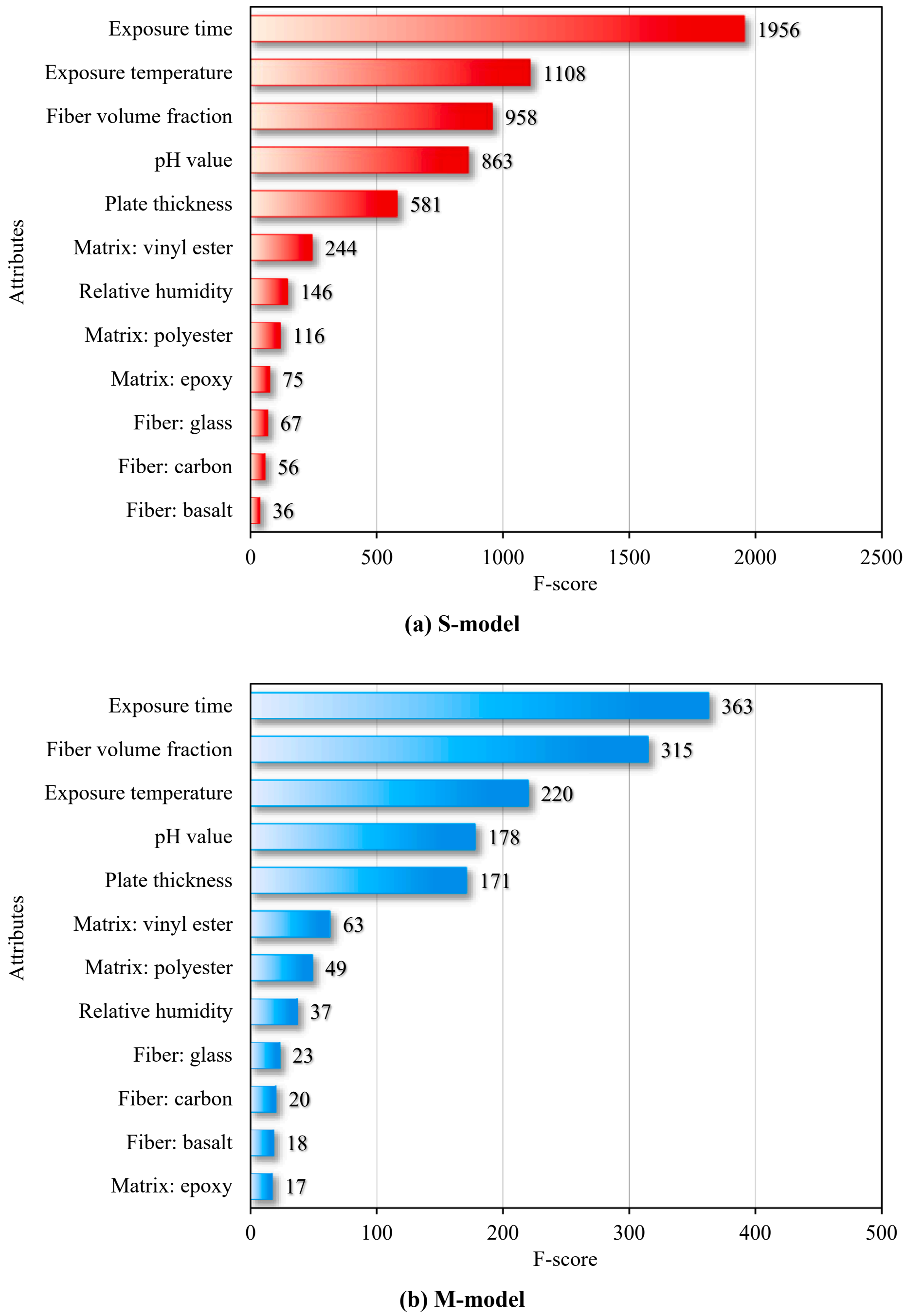

Fig. 9. Attribute importance analyses of XGBoost decision tree model: (a) For S-model, the exposure time and temperature, and pH value of environment are the most important external factors, and the fiber volume fraction and plate thickness are the most important internal factors. (b) For M−model, the most important factors are the same with those of S-model, except that the fiber volume fraction is shown to be more important than exposure temperature in determining the residual tensile modulus.

R^2 value is used as the primary evaluation criterion for assessing the prediction accuracy of predictive models. First, for S-model for the tensile strength, R^2 values of training set and testing set are all greater than 0.90, thus showing an excellent correlation between XGBoost predictions and experimental results. In fact, such a high prediction accuracy ($R^2 = 0.93$ in this case) is the highest in the available literature, particularly when the prediction model is to be cross-validated using a great number of test results from many studies. In addition, for M−model for the tensile modulus of elasticity, R^2 value of training set is 0.90, while that value of testing set is 0.85. Such a discrepancy between the training set and testing set is also reflected in Fig. 8. The main reason for this lower R^2 of testing set is that a much smaller database is available for developing the M−model. The volume of M−model dataset is 204, while that number of S-model is 542. Despite having a lower R^2 value than S-model, the prediction accuracy of M−model ($R^2 = 0.85$ in this case) is still sufficiently acceptable, particularly when compared to the other existing predictions.

In addition to R^2 value, RMSE, MAE and MAPE are used to conduct a secondary assessment with regard to the proposed XGBoost model. For S-model for tensile strength, all those errors are less than 0.07, and for M−model for tensile modulus, all those errors are less than 0.05. Such a small error (greatest error = 0.07, in this case) sufficiently demonstrates the excellent accuracy of both the S-model and M−model.

5.2. Attribute importance analysis

A number of factors may affect the residual mechanical properties of pultruded FRP composites subjected to environmental effects of water, high humidity and alkaline solution immersion. The possible influential factors may include the exposure time, exposure temperature, pH value,

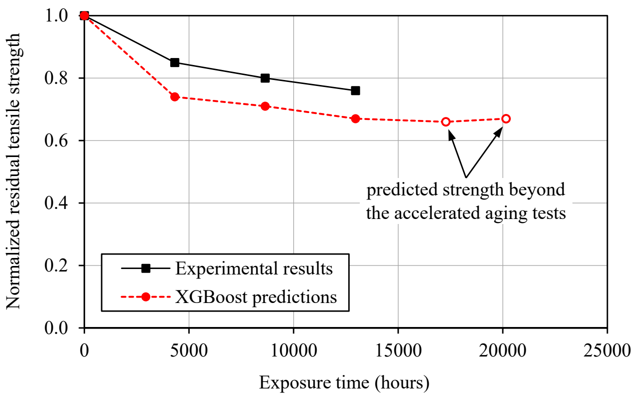

Fig. 10. Alkaline aging test results (strength) and XGBoost S-model predictions.

plate thickness, fiber type, matrix type and fiber volume fraction. All these factors, together, have a combined effect on the mechanical properties of FRP composites, and their synergistic effect makes prediction a challenging task for the field [10]. In this regard, XGBoost decision tree is adopted to analyze the attribute importance of each influential factor.

XGBoost model is known to have a good interpretability in terms of the output results. All attributes pertaining to the prediction results can be quantitatively described using an F-score. F-score indicates the contribution of an attribute in XGBoost decision tree. For instance, the F-score of a particular attribute is to increase when this attribute is increasingly used in the tree decision process. The F-scores of all attributes being investigated are calculated and presented in Fig. 9. The exposure time has the highest F-score for both the S-model and M−model. Thus, exposure time is the most influential factor of FRP degradation. Following exposure time, the exposure temperature and fiber volume fraction are the second and the third influential factors for S-model, while for M−model, the fiber volume fraction has a slightly higher F-score over exposure temperature. Then, for both the S-model and M−model, the pH value of environment and the plate thickness have a relatively significant effect on FRP composites. Finally, the fiber type and matrix type are observed to have the least impact on the residual mechanical properties of FRP composites.

From attribute importance analysis, it can be confidently concluded that the external factors of exposure time, exposure temperature and pH value of environment and the internal factors of fiber volume fraction and plate thickness are the most important factors for determining the residual tensile properties of pultruded FRP composites subjected to water, high humidity and alkaline solution immersion. In particular, exposure time is known as the most influential factor in the FRP degradation process. In this regard, it has been recommended for future work (see corresponding section in [50]) that the exposure time shall be from one-and-a-half years to three years so as to capture the real degradation process of FRP composites. In addition, fiber volume fraction and plate thickness are the inherent material properties that have a great impact on FRP degradation. Indeed, these two factors directly impact the degradation mechanisms of FRP composites. For instance, the degradation mechanisms of FRP composites immersed in water are shown in Fig. 1. The fiber content as well as the corresponding matrix content are identified to affect the degradation at fiber, matrix and interface levels. In conclusion, the findings from attribute importance analysis agree well with those reported in the literature.

6. Model robustness tests

6.1. Predictions of alkaline aging tests

In this work, an independent dataset of long-term properties of pultruded FRPs subjected to alkaline conditions from Al-Salloum et al. [3] was selected to test the robustness of proposed XGBoost model. The pH value of environment is 12.8 and the exposure temperature is 50 °C. The specimens were FRP bars with diameter of 12 mm and made of glass fiber and vinyl ester resin with a fiber volume fraction of 83%. The test was conducted by up to 12,960 h. Note that this dataset was not used in the training process. The experimentally determined residual tensile strength and the corresponding XGBoost predictions are presented in Fig. 10.

From Fig. 10, it can be seen that S-model could provide uniformly conservative predictions of residual tensile strength, and in those cases (exposure time = 4320, 8640, 12,960 h), the absolute differences between experimental results and model predictions are within 13%. In addition, using proposed S-model, the future trend of residual tensile strength can be predicted, as shown by the two blank dots in Fig. 10. These two dots represent the residual tensile strength measured at 17,280 and 20,160 h. Note that 20,160 h is the highest exposure time in training set.

6.2. Predictions of acidic aging tests

In the proposed XGBoost model, environmental conditions of water/high humidity and alkaline solution are differentiated by pH value, and these conditions with pH value less than 7 are automatically assumed to be the water condition. This assumption was also checked through a model robustness test. Due to the lack of acidic aging tests, only the test data (residual tensile strength, in this case) from Gentry et al. [34] was adopted in this work. The specimens were FRP plates with thickness of 6.35 mm and made of glass fiber and vinyl ester resin with a fiber volume fraction of 22%. All the test information was input into the S-model and the predicted residual tensile strength were obtained, as shown in Table 5.

Considering the proposed XGBoost model was not trained by any acidic aging test data, the absolute difference between experimental results and model predictions actually indicates the degree of degradation effect of acidic solution as compared to water. When pH value is 5 (i.e., weak acidic environment) under room temperature (23 °C, in this case), the greatest absolute difference between experiments and predictions is within 4%; hence, the aging effect of weak acidic solution is

Table 5
Acidic aging test results (strength) and XGBoost S-model predictions.

Author	Specimen	pH value	Exposure temperature (°C)	Exposure time (hours)	Experimentally-determined residual tensile strength (MPa)	XGBoost-predicted residual tensile strength (MPa)	pred/exp
Gentry et al. [34]	Control	7	23	0	403	–	–
	1	5	23	672	404	392	0.97
	2			2016	359	371	1.03
	3			5375	379	363	0.96
	4	3	23	168	392	416	1.06
	5			336	391	407	1.04
	6	3	80	168	307	319	1.04
	7			336	305	313	1.03
	8			672	242	278	1.15

Table A1
Residual tensile strength of pultruded FRP composites subjected to water and high humidity.

Author	Fiber/ Matrix	V_f	Coupon thickness (mm)	Coupon description	Aging effect	Exposure time (hours)	Exposure temperature (°C)	Experimentally-determined residual tensile strength (MPa)	XGBoost-predicted residual tensile strength (MPa)	pred/ exp
Shao and Kouadio [70]	G/P	0.58	4.7	top flange of sheet pile panel	control	0	23	433	433	1.00
Gentry et al. [34]	G/V	0.22	6.35	plate	deionized water	5376	23	394	362	0.92
						672	50	359	355	0.99
						2016	50	350	314	0.89
						168	80	302	318	1.05
						336	80	303	315	1.04
McBagonluri et al. [57]	G/V	0.29	3.175	plate	control	0	23	212	212	1.00
Chen et al. [19]	G/V	0.51	9.53 diameter	rebar	tap water	2880	23	732	747	1.02
Chu and Karbhari [25]	G/V	0.62	1.6	wet plate	deionized water	12,600	23	602	695	1.15
				dry plate	deionized water	1680	23	896	850	0.95
						3360	23	825	816	0.99
						1680	40	762	688	0.91
						3360	40	650	631	0.97
						12,600	40	533	496	0.93
						5040	60	521	542	1.04
						8400	60	449	422	0.94
Kafodya et al. [42]	C/E	–	1.4	plate	distilled water 0% strain	672	23	2080	2166	1.04
Lu et al. [51]	B/E	0.71	1.4	plate	distilled water	2160	20	1198	1167	0.97
						336	40	1471	1410	0.96
						1440	40	1226	1195	0.98
				135 °C aged plate	control	2160	20	980	949	0.98
					distilled water	336	40	1452	1392	0.96
				300 °C aged plate	distilled water	720	40	1060	1075	1.02
						1440	60	603	709	1.17
						2160	60	446	505	1.15
Grammatikos et al. [36]	G/P	0.45	6.4	plate	distilled water	1344	25	373	392	1.05
						1344	80	353	314	0.89
						5376	80	326	318	0.97
Heshmati et al. [39]	C/E	0.69	1.25	plate	distilled water	5040	20	2690	2660	0.99
					95% humidity distilled water	20,160	45	1447	1748	1.21
	G/P	0.58	10	plate	distilled water	5040	20	213	208	0.98
						20,160	20	166	189	1.14
					95% humidity	20,160	45	198	152	0.77
Kim et al. [44]	G/V	0.55	12.7	rod	tap water	2160	25	554	596	1.07
						720	40	619	640	1.04
						3168	80	541	527	0.97
	G/ modified V	0.50	12.7 diameter	rod	control	0	25	661	654	0.99
					tap water	1440	25	500	559	1.12
						3168	25	580	547	0.95
						720	80	439	432	0.99
						1440	80	365	365	0.99
	G/V	0.30		strand	distilled water	48	20	2748	2802	1.02
						480	20	2781	2646	0.95
						24	80	2129	2021	0.95
Cabral-Fonseca et al. [14]	G/P	0.68	5	box-section	demineralized water	6480	20	333	321	0.96
						4320	40	328	312	0.96
						12,960	40	340	328	0.95
						2160	60	326	318	1.00
	G/V	0.69	5	box-section	control	0	20	393	405	1.03
					demineralized water	4320	20	382	378	0.99
						2160	60	400	353	0.86
						4320	60	305	313	1.00
						8640	60	297	285	0.98
						4320	40	364	352	0.97
Zhang [83]	G/V	0.7	3	plate	deionized water	168	80	591	506	0.88

Table A2
Residual tensile strength of pultruded FRP composites subjected to alkaline solution.

Author	Fiber/ Matrix	V_f	Coupon thickness (mm)	Coupon description	Aging effect	Exposure time (hours)	Exposure temperature (°C)	Experimentally-determined residual tensile strength (MPa)	XGBoost-predicted residual tensile strength (MPa)	pred/ exp
Gentry et al. [34]	G/V	0.22	6.35	plate		336	80	283	291	1.03
Chu et al. [24]	G/V	0.62	1.6	plate	alkaline solution(pH = 11.5)	8400	40	539	502	0.94
						2520	60	473	482	1.02
						5040	60	450	432	0.97
						8400	60	396	405	1.02
						8400	80	364	364	1.01
Micelli and Nanni [58]	G/P		6.35 diameter	rod	control	0	23	2401	2329	0.97
Chen et al. [19]	G/V	0.51	9.53 diameter	bar		2880	23	751	751	1.00
Kafodya et al. [42]	C/E	–	1.4	plate	Seawater 30% strain	336	23	1870	1934	1.03
						2016	23	2200	2265	1.03
						3360	23	2170	2084	0.96
Lu et al. [51]	B/E	0.71	1.4	plate	alkaline solution (pH = 12.8)	720	20	1327	1312	0.99
						2160	20	1107	1046	0.95
						336	40	1407	1422	1.01
						720	60	1128	1236	1.09
						1440	60	1001	1047	1.05
				135 °C aged plate	control alkaline solution (pH = 12.8)	720	60	1081	1081	0.99
Heshmati et al. [39]	C/E	0.69	1.25	plate	saltwater	20,160	20	2319	2620	1.14
						20,160	45	1155	2006	1.74
Chen et al. [18]	G/V	0.51	9.53 diameter	rebar	alkaline solution (pH = 13.6)	2160	60	408	427	1.05
Kim et al. [44]	G/V	0.55	12.7 diameter	rod	seawater (3% NaCl)	1440	25	565	593	1.05
						3168	25	599	613	1.03
						720	80	621	580	0.93
						1440	80	558	530	0.96
					alkaline solution (pH = 13)	720	25	523	572	1.10
	G/ modified V	0.50	12.7 diameter	rod	seawater (3% NaCl)	720	40	592	599	1.02
						2160	40	538	525	0.98
						720	80	431	438	1.01
					alkaline solution (pH = 13)	1440	25	547	488	0.90
						1440	40	539	473	0.87
						1800	40	532	466	0.87
						1440	80	327	334	1.03
	G/V	0.30	0.7 diameter	strand	alkaline solution (pH = 13)	2880	20	519	628	1.23
						720	80	176	126	0.69
						1440	80	140	168	1.25
						3600	80	82	55	0.60
					seawater (3% NaCl)	240	20	2717	2529	0.93
					seawater (4% $CaCl_2$)	720	20	2376	2376	1.00
Won et al. [76]	G/V	–	12.7 diameter	rebar	alkaline solution	4320	20	600	621	1.04
						720	40	677	670	0.99
						7200	40	483	531	1.11
						720	80	564	550	0.98
						5760	80	482	461	0.95
Sawpan et al. [69]	G/E	–	14 diameter	rebar	concrete pore solution	720	60	860	878	1.02

(continued on next page)

Table A2 (*continued*)

Author	Fiber/ Matrix	V_f	Coupon thickness (mm)	Coupon description	Aging effect	Exposure time (hours)	Exposure temperature (°C)	Experimentally-determined residual tensile strength (MPa)	XGBoost-predicted residual tensile strength (MPa)	pred/ exp
Cabral-Fonseca et al. [14]	G/P	0.68	5	box-section	saltwater (35 g/L NaCl)	8640	20	348	368	1.06
						2160	40	419	382	0.92
						12,960	60	312	288	0.92
	G/V	0.69	5	box-section	saltwater (35 g/L NaCl)	6480	20	317	356	1.12
						8640	20	373	397	1.06
						2160	40	393	401	1.02
						8640	40	354	354	0.99
						2160	60	373	357	0.96
						4320	60	322	326	1.01

similar to that of water, and the model is capable of providing accurate predictions. When pH value decreases to 3 (i.e., strong acidic environment), the absolute differences between experimental results and XGBoost predictions are still within 6% for those tests conducted by up to 336 h at 80 °C. Then, XGBoost prediction starts to differ from experimental result when FRP material was immersed in a strong acidic solution (pH = 3) at 80 °C by up to 672 h. This is consistent with those findings reported by Amaro et al. [4] and Feng et al. [32] that acidic solution has a great impact on mechanical properties of FRP composites only when the materials experienced a long aging period at a high temperature. It is noted that this model can be further improved when acidic aging test data is augmented.

7. Conclusions

In this work, an XGBoost decision tree-based predictive model was developed for calculating the residual tensile strength and modulus of pultruded FRP composites exposed to water, high humidity and alkaline

Table A3
Residual tensile modulus of pultruded FRP composites subjected to water and high humidity.

Author	Fiber/ Matrix	V_f	Coupon thickness (mm)	Coupon description	Aging effect	Exposure time (hours)	Exposure temperature (°C)	Experimentally-determined residual tensile modulus (MPa)	XGBoost-predicted residual tensile modulus (MPa)	pred/ exp
Shao and Kouadio [70]	G/P	0.58	4.7	top flange of sheet pile panel	distilled water	6240	23	32,300	34,094	1.05
		0.37	3.2	web of sheet pile panel	control	0	23	11,700	12,753	1.09
					distilled water	504	100	13,800	11,578	0.84
	C/E	–	1.4	plate	distilled water 0% strain	336	23	167,421	162,770	0.98
					distilled water 50% strain	3360	23	161,324	165,978	1.03
Lu et al. [51]	B/E	0.71	1.4	plate	control	0	23	59,400	58,212	0.98
					distilled water	2160	40	52,480	52,480	1.00
				135 °C aged plate		2160	60	49,220	51,089	1.04
				300 °C aged plate		2160	20	54,370	53,162	0.98
						2160	40	48,080	51,732	1.07
Grammatikos et al. [36]	G/P	0.45	6.4	plate	distilled water	672	60	25,000	24,123	0.96
						2688	60	25,000	25,000	1.00
						5376	80	27,000	25,244	0.94
Heshmati et al. [39]	C/E	0.69	1.25	plate	distilled water	5040	20	154,163	149,538	0.97
						20,160	45	151,404	156,039	1.03
Kim et al. [44]	G/V	0.55	12.7 diameter	rod	tap water	720	25	33,123	34,278	1.04
						2160	25	29,834	30,996	1.04
						2160	40	35,002	31,156	0.89
						3168	40	28,738	28,355	0.99
	G/ modified V	0.50	12.7 diameter	rod	tap water	1440	80	38,821	39,248	1.01
						2160	80	39,950	38,675	0.96
						3168	80	35,086	33,374	0.95
Al-Salloum et al. [3]	G/V	0.83	12 diameter	rebar	tap water	4320	23	58,900	56,496	0.96
Zhang [83]	G/V	0.7	3	plate	deionized water	168	80	23,560	24,246	1.03

Table A4
Residual tensile modulus of pultruded FRP composites subjected to alkaline solution.

Author	Fiber/ Matrix	V_f	Coupon thickness (mm)	Coupon description	Aging effect	Exposure time (hours)	Exposure temperature (°C)	Experimentally-determined residual tensile modulus (MPa)	XGBoost-predicted residual tensile modulus (MPa)	pred/ exp
Micelli and Nanni [58]	C/E-V modified		8.26 diameter	rod	alkaline solution (pH = 13.0)	1008	60	30,426	31,619	1.04
Kafodya et al. [42]	C/E	–	1.4	plate	seawater 0% strain	2016	23	167,381	162,732	0.98
					seawater 50% strain	3360	23	173,452	170,355	0.99
Lu et al. [51]	B/E	0.71	1.4	135 °C aged plate	alkaline solution (pH = 12.8)	2160	20	51,870	51,870	1.00
Heshmati et al. [39]	G/P	0.58	10	plate	saltwater	20,160	20	13,502	13,080	0.97
						5040	45	12,811	12,672	0.99
Kim et al. [44]	G/V	0.55	12.7 diameter	rod	seawater (3% NaCl)	720	25	35,094	33,551	0.96
						1440	40	28,776	29,927	1.04
						3168	40	28,608	28,608	1.00
						720	80	31,152	31,921	1.02
					alkaline solution (pH = 13)	720	80	35,853	37,395	1.05
	G/ modified V	0.50	12.7 diameter	rod	seawater (3% NaCl)	1440	40	38,683	38,258	0.98
						2160	40	36,613	36,613	1.00
						720	80	41,060	38,921	0.95
						3168	80	34,248	33,820	0.98
Sawpan et al. [69]	G/E	–	14 diameter	rebar	concrete pore solution	1440	60	54,760	54,760	1.00
						2880	60	54,350	55,997	1.03

solution. XGBoost decision tree, as a machine learning technique, is capable of solving the nonlinear regression problems in various engineering fields. To develop the XGBoost model, the FRP aging database created by Liu et al. [50] was used, and for the scope of this work, a total of 746 test data were collected, including 542 tensile strength data and 204 tensile modulus data. In addition, detailed methodology of XGBoost decision tree was presented. The model predictions were found to have an excellent agreement with experimental results. Following conclusions are drawn from this work.

(1) A novel and effective approach for quantitatively predicting the durability based on the test data and the algorithm is presented. The machine learning analysis/evaluation framework—XGBoost decision tree—is demonstrated to be effective in predicting the residual tensile properties of pultruded FRP composites exposed to water, high humidity, and alkaline solution. The predictions were validated through a dataset with 108 tensile strength and 41 tensile moduli. The R^2 values for predicted tensile strength and moduli are 0.93 and 0.85, respectively. Having such an amount of cross-validations, the observed prediction accuracy of proposed XGBoost model is the highest in available literature.
(2) The XGBoost decision tree model is able to provide a good prediction interpretability. All the attributes of input data can be quantitatively analyzed regarding their importance, including the exposure time, exposure temperature, pH value of environment, fiber volume fraction, plate thickness, fiber type and matrix type. The exposure time and temperature were found to be the most important external factors, and the fiber volume fraction and plate thickness were the most important internal factors. Attribute importance analysis could readily serve as a guide for future research.
(3) The proposed XGBoost model provides a new approach for solving the conventional engineering problems, particularly for those problems involving a number of influential factors. Additionally, this model can be continuously updated when necessary dataset is further enriched. This work only focuses on the tensile properties of FRP composites and only three environmental effects are considered. In the future, this XGBoost model can be updated to predict the compressive, shear and flexural properties of FRP composites subjected to various environmental effects.

8. Data availability

All data, models, and code generated or used during the study are available at the corresponding author.

CRediT authorship contribution statement

Xing Liu: Conceptualization, Methodology, Software, Investigation, Formal analysis, Data curation, Visualization, Writing – original draft. **TianQiao Liu:** Methodology, Investigation, Visualization, Writing – original draft. **Peng Feng:** Conceptualization, Methodology, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

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

Acknowledgements

This research was supported by the National Key R&D Program of China (2018YFB1501200), the National Natural Science Foundation of China (No. U2106219), and the grants from the Institute for Guo Qiang, Tsinghua University (2019GQG1004 and 2020GQC0004).

Appendix A. XGBoost-predicted residual tensile properties of pultruded FRP composites subjected to water, high humidity and alkaline solution

In this section, XGBoost-predicted residual tensile properties as well as the detailed information regarding the aging tests are presented. Note that in Tables A1–A4, only the data from testing set is presented. The fiber and matrix types are designated following this rule: G for glass fiber, C for carbon fiber, V for vinyl ester resin, P for polyester resin and E for epoxy resin. For instance, the fiber/matrix composition denoted as G/V indicates the glass fiber reinforced vinyl ester resin matrix.

References

[1] Afshar A, Liao H-T, Chiang F-P, Korach CS. Time-dependent changes in mechanical properties of carbon fiber vinyl ester composites exposed to marine environments. Compos Struct 2016;144:80–5.
[2] Akay M, Ah Mun SK, Stanley A. Influence of moisture on the thermal and mechanical properties of autoclaved and oven-cured Kevlar-49/epoxy laminates. Compos Sci Technol 1997;57(5):565–71.
[3] Al-Salloum YA, El-Gamal S, Almusallam TH, Alsayed SH, Aqel M. Effect of harsh environmental conditions on the tensile properties of GFRP bars. Compos B Eng 2013;45(1):835–44.
[4] Amaro AM, Reis PNB, Neto MA, Louro C. Effects of alkaline and acid solutions on glass/epoxy composites. Polym Degrad Stab 2013;98(4):853–62.
[5] Arrhenius S. XXXI. On the influence of carbonic acid in the air upon the temperature of the ground. Lond Edinb Dub Philosoph Mag J Sci 1896;41(251):237–76.
[6] American Society of Civil Engineers. Pre-Standard for Load and Resistance Factor Design (LRFD) of Pultruded Fiber Reinforced Polymer (FRP) Structures, ASCE, 2010.
[7] Ascione L, Caron JF, Godonou P, van IJselmuijden K, Knippers J, Mottram T, et al. Prospect for new guidance in the design of FRP (EUR27666). Publications Office of the European Union; 2016.
[8] Baltzis D, Bekas D, Tsirka K, Parlamas A, Ntaflos A, Zafeiropoulos N, et al. Multi-scaled carbon epoxy composites underwater immersion: A durability study. Compos Sci Technol 2020;199:108373. https://doi.org/10.1016/j.compscitech.2020.108373.
[9] Bazli M, Ashrafi H, Oskouei AV. Effect of harsh environments on mechanical properties of GFRP pultruded profiles. Compos Part B-Eng 2016;99:203–15.
[10] Benzarti K, Colin X. Understanding the durability of advanced fibre-reinforced polymer (FRP) composites for structural applications. In *Advanced Fibre-Reinforced Polymer (FRP) Composites for Structural Applications*. In: Advanced Fibre-Reinforced Polymer (FRP) Composites for Structural Applications. Elsevier; 2013. p. 361–439. https://doi.org/10.1533/9780857098641.3.361.
[11] Biredagn NK, Nehemiah HK, Kannan A. Knowledge Mining from Clinical Datasets Using Rough Sets and Backpropagation Neural Network. Comput Math Meth Med 2015;2015:460189.
[12] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, et al., API design for machine learning software: experiences from the scikit-learn project, arXiv preprint arXiv:1309.0238, 2013.
[13] Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature 2018;559(7715):547–55.
[14] Cabral-Fonseca S, Correia JR, Rodrigues MP, Branco FA. Artificial accelerated ageing of GFRP pultruded profiles made of polyester and vinylester resins: characterisation of physical–chemical and mechanical damage. Strain 2012;48(2):162–73.
[15] Cascardi A, Micelli F, Aiello MA. An Artificial Neural Networks model for the prediction of the compressive strength of FRP-confined concrete circular columns. Eng Struct 2017;140:199–208.
[16] Chakraborty D, Elzarka H. Advanced machine learning techniques for building performance simulation: A comparative analysis. J Build Perform Simul 2019;12(2):193–207.
[17] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016.
[18] Chen Yi, Davalos JF, Ray I. Durability Prediction for GFRP Reinforcing Bars Using Short-Term Data of Accelerated Aging Tests. J Compos Constr 2006;10(4):279–86.
[19] Chen Yi, Davalos JF, Ray I, Kim H-Y. Accelerated aging tests for evaluations of durability performance of FRP reinforcing bars for concrete structures. Compos Struct 2007;78(1):101–11.
[20] Cheng M-Y, Chou J-S, Roy AFV, Wu Y-W. High-performance concrete compressive strength prediction using time-weighted evolutionary fuzzy support vector machines inference model. Autom Constr 2012;28:106–15.
[21] China Association for Engineering Construction Standardization, Technical specification for pultruded fiber reinforced polymer composites structure, T/CECS 692-2020 (In Chinese).
[22] Chou J-S, Chiu C-K, Fahmoud M, AI-Taharwa I. Optimizing the Prediction Accuracy of Concrete Compressive Strength Based on a Comparison of Data-Mining Techniques. J Comput Civ Eng 2012;30(4):1–11.
[23] Chou J-S, Tsai C-F, Pham A-D, Lu Y-H. Machine learning in concrete strength simulations: Multi-nation data analytics. Constr Build Mater 2014;73:771–80.
[24] Chu W, Wu L, Karbhari VM. Durability evaluation of moderate temperature cured E-glass/vinylester systems. Compos Struct 2004;66(1-4):367–76.
[25] Chu W, Karbhari VM. Effect of water sorption on performance of pultruded E-glass/vinylester composites. J Mater Civ Eng 2005;17(1):63–71.
[26] Cromwell JR, Harries KA, Shahrooz BM. Environmental durability of externally bonded FRP materials intended for repair of concrete structures. Constr Build Mater 2011;25(5):2528–39.
[27] Davalos JF, Chen Yi, Ray I. Long-term durability prediction models for GFRP bars in concrete environment. J Compos Mater 2012;46(16):1899–914.
[28] Davies P, MazÉas F, Casari P. Sea Water Aging ofGlass Reinforced Composites: Shear Behaviour andDamage Modelling. J Compos Mater 2001;35(15):1343–72.
[29] Dettling M, Buhlmann P. Boosting for tumor classification with gene expression data. Bioinformatics 2003;19(9):1061–9.
[30] Dong W, Huang Y, Lehane B, Ma G. XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring. Autom Constr 2020;114:103155. https://doi.org/10.1016/j.autcon.2020.103155.
[31] Duan J, Asteris PG, Nguyen H, Bui X-N, Moayedi H. A novel artificial intelligence technique to predict compressive strength of recycled aggregate concrete using ICA-XGBoost model. Adv Neural Inform Process Syst 2020.
[32] Feng P, Wang J, Tian Y, Loughery D, Wang Y. Mechanical Behavior and Design of FRP Structural Members at High and Low Service Temperatures. J Compos Constr 2016;20(5):04016021. https://doi.org/10.1061/(ASCE)CC.1943-5614.0000676.
[33] Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. Ann Stat 2001;29(5):1189–232.
[34] Gentry TR, Bank LC, Barkatt A, Prian L. Accelerated test methods to determine the long-term behavior of composite highway structures subject to environmental loading. J Compos Technol Res 1998;20(1):38–50.
[35] Safarzadegan Gilan S, Bahrami Jovein H, Ramezanianpour AA. Hybrid support vector regression–Particle swarm optimization for prediction of compressive strength and RCPT of concretes containing metakaolin. Constr Build Mater 2012;34:321–9.
[36] Grammatikos SA, Evernden M, Mitchels J, Zafari B, Mottram JT, Papanicolaou GC. On the response to hygrothermal aging of pultruded FRPs used in the civil engineering sector. Mater Des 2016;96:283–95.
[37] A. Graves, Adaptive Computation Time for Recurrent Neural Networks, 2016, pp. 1–19.
[38] He X, Chua TS. Neural factorization machines for sparse predictive analytics. In: SIGIR 2017 - Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval; 2017. p. 355–64.
[39] Heshmati M, Haghani R, Al-Emrani M. Durability of bonded FRP-to-steel joints: Effects of moisture, de-icing salt solution, temperature and FRP type. Compos B Eng 2017;119:153–67.
[40] Hollaway LC. A review of the present and future utilisation of FRP composites in the civil infrastructure with reference to their important in-service properties. Constr Build Mater 2010;24(12):2419–45.
[41] Jiang K, Han Q, Bai Y, Du X. Data-driven ultimate conditions prediction and stress-strain model for FRP-confined concrete. Compos Struct 2020;242:112094. https://doi.org/10.1016/j.compstruct.2020.112094.
[42] Kafodya I, Xian G, Li H. Durability study of pultruded CFRP plates immersed in water and seawater under sustained bending: Water uptake and effects on the mechanical properties. Compos B Eng 2015;70:138–48.
[43] Kim B, Khanna R, Koyejo O. Examples are not enough, learn to criticize! Criticism for interpretability. Adv Neural Inform Process Syst 2016:2288–96 (Nips).
[44] Kim H-Y, Park Y-H, You Y-J, Moon C-K. Short-term durability test for GFRP rods under various environmental conditions. Compos Struct 2008;83(1):37–47.
[45] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Ijcai 1995;14(2):1137–45.
[46] Li Z, Furmanski J, Lepech MD. Micromechanics modeling and homogenization of glass fiber reinforced polymer composites subject to synergistic deterioration. Compos Sci Technol 2021;203:108629. https://doi.org/10.1016/j.compscitech.2020.108629.
[47] Liao K, Schultheisz CR, Hunston DL. Effects of environmental aging on the properties of pultruded GFRP. Compos B Eng 1999;30(5):485–93.
[48] Lim S, Chi S. Xgboost application on bridge management systems for proactive damage estimation. Adv Eng Inf 2019;41:100922. https://doi.org/10.1016/j.aei.2019.100922.
[49] Liu TQ, Feng P, Wu Y, Liao S, Meng X. Developing an innovative curved-pultruded large-scale GFRP arch beam. Compos Struct 2021;256:113111. https://doi.org/10.1016/j.compstruct.2020.113111.
[50] Liu TQ, Liu X, Feng P. A comprehensive review on mechanical properties of pultruded FRP composites subjected to long-term environmental effects. Compos B Eng 2020;191(March):107958.
[51] Lu Z, Xian G, Li H. Effects of exposure to elevated temperatures and subsequent immersion in water or alkaline solution on the mechanical properties of pultruded BFRP plates. Compos B Eng 2015;77:421–30.
[52] Ma X, Sha J, Wang D, Yu Y, Yang Q, Niu X. Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning. Electron Commer Res Appl 2018;31:24–39.
[53] Ma Y, Jin S, Yokozeki T, Ueda M, Yang Y, Elbadry EA, et al. Effect of hot water on the mechanical performance of unidirectional carbon fiber-reinforced nylon 6 composites. Compos Sci Technol 2020;200:108426. https://doi.org/10.1016/j.compscitech.2020.108426.
[54] Mansouri I, Ozbakkaloglu T, Kisi O, Xie T. Predicting behavior of FRP-confined concrete using neuro fuzzy, neural network, multivariate adaptive regression splines and M5 model tree techniques. Mater Struct 2016;49(10):4319–34.

[55] MarchLeuba, Jose A, Huang, Tai L. A Review of the Physical Phenomena that Impact the Stability of BWRs. Nber Chapters 2008:131–67.

[56] Marks M, Jóźwiak-Niedźwiedzka D, Glinicki MA, Olek J, Marks M. Assessment of scaling durability of concrete with CFBC ash by automatic classification rules. J Mater Civ Eng 2012;24(7):860–7.

[57] McBagonluri F, Garcia K, Hayes M, Verghese KNE, Lesko JJ. Characterization of fatigue and combined environment on durability performance of glass/vinyl ester composite for infrastructure applications. Int J Fatigue 2000;22(1):53–64.

[58] Micelli F, Nanni A. Durability of FRP rods for concrete structures. Constr Build Mater 2004;18(7):491–503.

[59] Miller T. Explanation in Artificial Intelligence: Insights from the Social Sciences. Artif Intell 2017;267.

[61] Mozumder RA, Roy B, Laskar AI. Support vector regression approach to predict the strength of FRP confined concrete. Arab J Sci Eng 2017;42(3):1129–46.

[62] Naya F, Herráez M, Lopes CS, González C, Van der Veen S, Pons F. Computational micromechanics of fiber kinking in unidirectional FRP under different environmental conditions. Compos Sci Technol 2017;144:26–35.

[63] Nielsen D. Tree boosting with xgboost-why does xgboost win“ every” machine learning competition? NTNU; 2016. Master's thesis.

[64] Park C, Padgett WJ. Stochastic degradation models with several accelerating variables. IEEE Trans Reliab 2006;55(2):379–90.

[65] Perera R, Arteaga A, Diego AD. Artificial intelligence techniques for prediction of the capacity of RC beams strengthened in shear with external FRP reinforcement. Compos Struct 2010;92(5):1169–75.

[66] Perera R, Tarazona D, Ruiz A, Martín A. Application of artificial intelligence techniques to predict the performance of RC beams shear strengthened with NSM FRP rods. Formulation of design equations. Compos B Eng 2014;66:162–73.

[67] Pham TM, Hadi MNS. Predicting stress and strain of FRP-confined square/rectangular columns using artificial neural networks. J Compos Constr 2014;18(6):1–9.

[68] Ranjan GSK, Verma AK, Radhika S. K-nearest neighbors and grid search cv based real time fault monitoring system for industries. In: In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*; 2019. p. 1–5.

[69] Sawpan MA, Mamun AA, Holdsworth PG. Long term durability of pultruded polymer composite rebar in concrete environment. Mater Des 2014;57:616–24.

[70] Shao Y, Kouadio S. Durability of fiberglass composite sheet piles in water. J Compos Constr 2002;6(4):280–7.

[71] Taffese WZ, Sistonen E. Machine learning for durability and service-life assessment of reinforced concrete structures: Recent advances and future directions. Autom Constr 2017;77:1–14.

[72] Tam L-h, Zhou Ao, Wu C. Nanomechanical behavior of carbon fiber/epoxy interface in hygrothermal conditioning: A molecular dynamics study. Mater Today Commun 2019;19:495–505.

[73] Tixier A-J-P, Hallowell MR, Rajagopalan B, Bowman D. Application of machine learning to construction injury prediction. Autom Constr 2016;69(SEP.):102–14.

[74] Vedernikov A, Safonov A, Tucci F, Carlone P, Akhatov I. Pultruded materials and structures: A review. J Compos Mater 2020;54(26):4081–117.

[75] Wang W, Shi Y, Lyu G, Deng W. Electricity consumption prediction using xgboost based on discrete wavelet transform. DEStech Trans Comput Sci Eng 2017.

[76] Won J-P, Lee S-J, Kim Y-J, Jang C-I, Lee S-W. The effect of exposure to alkaline solution and water on the strength–porosity relationship of GFRP rebar. Compos B Eng 2008;39(5):764–72.

[77] Woo RSC, Chen Y, Zhu H, Li J, Kim J-K, Leung CKY. Environmental degradation of epoxy–organoclay nanocomposites due to UV exposure. Part I: Photo-degradation. Compos Sci Technol 2007;67(15-16):3448–56.

[78] Woo R, Zhu H, Leung C, Kim J. Environmental degradation of epoxy-organoclay nanocomposites due to UV exposure: Part II residual mechanical properties. Compos Sci Technol 2008;68(9):2149–55.

[79] Yeh I-C, Lien L-C. Knowledge Discovery Of Concrete Material Using Genetic Operation Trees. Expert Syst Appl 2009;36(3):5807–12.

[80] Yucong W, Bo W. Research on EA-Xgboost Hybrid Model for Building Energy Prediction. J. Phys. Conf. Series 2020;1518:12082.

[81] Zafar A, Bertocco F, Schjødt-Thomsen J, Rauhe JC. Investigation of the long term effects of moisture on carbon fibre and epoxy matrix composites. Compos Sci Technol 2012;72(6):656–66.

[82] Zhuowen Tu, Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering, in: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, vol. 2, 2005, pp. 1589–1596.

[83] Zhang X. Study on long-term performance of pultruded FRP for ocean applications. Tsinghua University; 2018. MS thesis.