Ph.D. Thesis

Development of interpretable AI to explore effective components of CAD/CAM resin composites

Osaka University Graduate School of Dentistry
Course for Oral Sciences
Department of Dental Biomaterials
Supervisor: Professor Satoshi Imazato

Hefei Li

I. Table of contents

I. Introduction

In the last 20 years, materials used for dental restoration have changed from alloys and metals to ceramics and resin composites because of the increasing aesthetic demand of patients. Especially, mechanical properties of resin composites have been greatly advanced based on numerous studies, and inlays and crowns made from resin composite blocks (RCBs) using computer-aided design and computer-aided manufacturing (CAD/CAM) systems are widely used in daily clinical practice [1]. However, despite the continued evolution of dental materials and manufacturing techniques, edge chipping and fracture of composite restorations still occur under clinical service, which results in the limited longevity of restorations [2, 3]. Therefore, how to improve the mechanical properties of resin composites has always been a topic of great interest, and efforts to improve the mechanical properties of CAD/CAM RCBs, including the flexural strength, have been made by changing the type of fillers and monomers [4-6]. However, because various compositions result in the varied flexural strength of CAD/CAM RCBs, which filler or monomer contributes specifically to improving flexural strength remains unknown.

The conventional approach to modifying material compositions to achieve superior properties relies on repetitive *in vitro* experiments, which are time-consuming and often inefficient [7], and can barely distinguish the contribution of each composition to the target material properties. In recent years, artificial intelligence (AI) technologies have become widely accepted in society, and preliminarily implemented in dentistry [8, 9]. A promising approach that combines traditional experimental methods with intelligent data analysis grew out of the quest for AI and is known as machine learning (ML). ML is a powerful

tool for finding meaningful regularities in high-dimensional data, which allows for predictions of unknown data. It uses algorithms by which a computer can learn from empirical data by modeling the linear or nonlinear relationships between material properties and related factors [7, 10]. ML has successfully resolved the difficulties of modeling the relationships between material properties and complex physical factors [11]. Compared with using only experimental measurement, ML quickly assesses and analyzes the collected data with extraction of various relevant features, which saves a great amount of time and cost for scientists and manufacturers [12]. Compared with conventional statistical modelling, ML methods are especially advantageous to deal with data that the number of input features surpasses the number of subjects [13]. On the other hand, ML methods yielded higher prediction accuracy when facing large dataset. ML makes minimal assumptions about the data-generating systems; they remain effective in scenarios where data collection lacks a carefully controlled experimental design and involves complex nonlinear interactions [13].

Moreover, for many AI users in both the dental and medical fields, a high level of accountability is required; thus, it is highly possible that directly interpretable and tractable AI techniques will be adopted as assistants for decision-making [14, 15]. In interpretable machine learning (IML), methods and models are proposed that make the behavior and predictions of machine learning systems understandable to humans [16]. Unlike the opaque explainability of deep neural networks, IML could provide insight into what a trained model has learned and reason about its entire decision-making process [17].

For the purpose of predicting material properties, ML models should contain three parts: training data, descriptors, and algorithms that can map the descriptors to the property of

interest [18]. The descriptors are a set of attributes that require selection to be capable of both uniquely defining each of the materials in the dataset and relating to the targeted property [18]. Compositional information has been commonly used as descriptors for developing ML models for materials [19]. Hyperparameters are parameters of ML algorithms that must be pre-set and tuned to control how an algorithm learns from the training data [20]. The development of an effective model also relies on the optimization of the hyperparameters [21].

To date, dental applications of ML, such as computer-aided diagnosis, treatment, and disease prediction, have mainly focused on classification problems for which the outputs are discrete values [8, 9, 22]. The correlation between the composition and mechanical behavior of aging enamel was considered in only one study [23]. There have been no studies in which ML has been applied to prediction of dental material properties. However, this data-driven informatic (materials informatics) approach [24] has already been applied to classical materials science to predict the mechanical properties (e.g., toughness, strength, and stiffness) of composite materials, and densities and elastic modulus of SiO_2-based glass using regression models with only the chemical composition or with the chemical or physical quantities as descriptors [25-28].

The aims of this study were:

1) to develop an interpretable AI using ML methods to predict the flexural strength of CAD/CAM RCBs, and
2) to investigate the components that affect flexural strength to explore the optimum composition based on the available dataset.

II. Machine learning model development

1. Purpose

In this chapter, development of ML models to predict the flexural strengths of CAD/CAM RCBs based on the available dataset was aimed. The fluctuation range and increased sample number were evaluated for dataset augmentation. Models were developed using the augmented dataset after hyperparameters adjustment and model evaluation.

2. Materials and methods

1) CAD/CAM RCBs

Twelve commercially available CAD/CAM RCBs were considered in this study: Cerasmart (CS; GC, Tokyo, Japan), Katana Avencia Block (KA; Kuraray Noritake Dental, Tokyo, Japan), Katana Avencia P Block (AP; Kuraray Noritake Dental), Shofu Block HC (HC; Shofu, Kyoto, Japan), Shofu Block HC Hard (HC; Shofu), KZR-CAD HR2 (HR2; Yamakin, Osaka, Japan), Estelite Block (EB; Tokuyama Dental, Tsukuba, Japan), Estelite P Block (EP; Tokuyama Dental), Brilliant Crios (BC; Coltene, Switzerland), Lava Ultimate (LU; 3M ESPE, St. Paul, MN, USA), Paradigm MZ 100 (MZ100; 3M ESPE), and Tetric CAD (TC; Ivoclar Vivadent, Schaan Liechtenstein). The details of the composition of each block are summarized in Table 1.

2) Data collection

The detailed compositional information of each commercial product was collected from manufacturers or the literatures [5, 29-33]. The flexural strengths of the products were collected from the manufacturers or literature, and they were measured by the three-point bending test according to the ISO 6872:2015 [34-36].

3) Investigation of fluctuation range

The initial data consisted of 16 attributes (15 input compositional descriptors and 1 label) and 12 samples. There were two types of compositional descriptors: fillers and monomers. The fillers were SiO_2, ZrO_2, $ZrSiO_4$, micro-fumed silica, barium glass, Al_2O_3, methacrylate mixed filler, SiO_2-ZrO_2 filler, and the monomers were urethane dimethacrylate (UDMA), 2,2-Bis(4-methacryloxypolyethoxyphenyl)propane (Bis-MEPP), triethylene glycol dimethacrylate (TEGDMA), neopentyl glycol dimethacrylate (NPGDMA), bisphenol A glycidyl methacrylate (Bis-GMA), and ethoxylated bisphenol A-glycol dimethacrylate (Bis-EMA) (Table 2). Filler contents were also added as one descriptor.

Each product had a unique composition; accordingly, 1 and 0 were used to represent whether the sample contained the specific descriptor or not, respectively. Therefore, the composition of each product could be represented with different combinations of 1 and 0. Considering that the input data for each sample were recognized as a multi-dimensional vector, a fluctuation range was defined for each vector. To decide the appropriate fluctuation range, different fluctuation ranges were set for 0 and 1 (Table 3).

Taking a fluctuation range 0.1 as an example, new samples (n = 9) were created based on the data of the original sample. Specifically, for each original sample, new samples were

constructed by generating random numbers to two decimal places using the function "RANDBETWEEN" in Microsoft Excel (version 2307) within the range [0, 0.1] for 0 and [0.95, 1.05] for 1. The same labels were used for the original samples and created samples (*i.e.*, flexural strength). Consequently, the dataset was increased to 16 attributes (15 input compositional descriptors and 1 label) and 120 samples. The same process was conducted for each of the fluctuation range; 10 different datasets were obtained accordingly.

4) Hyperparameters adjustment

Hyperparameters are parameters used to configure a ML algorithm and defines its architecture [19]. Four regression algorithms, random forest (RF) [37], extra trees (ET) [38], gradient boosting decision tree (GBDT) [39], and extreme gradient boosting (XGBoost) [40], were implemented using the scikit-learn package (version 0.24.1) in Python (version 3.7.4) and run in Jupyter Notebook (version 6.0.1) on a laptop (Surface Laptop 2: Core i5-8250U CPU and 8 GB RAM, Microsoft, Redmond, Washington, USA). Each of the dataset was imported and randomly split into two groups: 80% of the data was used for training the algorithm and 20% was used for testing [41]. The variable "random_state" was set to five different values (random_state= 1, 3, 9, 23, 100) to get different splits of each dataset. The function "GridSearchCV" in scikit-learn's model selection package was applied to search for the optimal combination of hyperparameters for each algorithm. This function evaluated the algorithm's performance for each combination of hyperparameters using 10-fold cross-validation, in which the algorithm fitted the training data 10 times (Figure 1). For each iteration, the training data were split into 10 subsets: 9 subsets were used for training the algorithm and the 10th subset was used

as test data. The combination of hyperparameters that exhibited the best performance during the cross-validation process was selected.

5) Algorithm performance on each fluctuation range dataset

The coefficient of determination (R^2), root mean square error (*RMSE*), and mean absolute error (*MAE*) were calculated to assess the regression accuracy of each algorithm on each fluctuation range dataset, accordingly ten R^2 values, ten *RMSE* values, and ten *MAE* values were obtained for each algorithm. The values of each metric were compared among different fluctuation range datasets numerically. The fluctuation range dataset showed the highest R^2 value, and the lowest *RMSE* value, *MAE* value was selected as the appropriate fluctuation range. The metrics are expressed as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)^2}{\sum_{i=1}^{m}\left(\bar{y} - y^{(i)}\right)^2} \quad (1)$$

$$RMSE = \frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - \hat{y}^{(i)}\right)^2 \quad (2)$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|y^{(i)} - \hat{y}^{(i)}\right|, \quad (3)$$

where $y^{(i)}$ is the flexural strength calculated from *in vitro* experiments, $\hat{y}^{(i)}$ is the predicted flexural strength from the above four algorithms, and m is the number of test samples.

6) Investigation of increased sample number

After the appropriate fluctuation range was settled, the appropriate increased sample number was investigated by generating different sample number datasets. Specifically, new samples (n = 1, 2, 3…9, 11) were constructed by generating random numbers to two decimal places within the decided fluctuation range [0, 0.1] for 0 and [0.95, 1.05] for 1 using the function "RANDBETWEEN" in Microsoft Excel (version 2307). The same labels were used for the original samples and created samples (*i.e.*, flexural strength). Accordingly, for each algorithm, eleven datasets with different increased sample numbers were obtained. After hyperparameters adjustment as described in section 4), R^2, *RMSE,* and *MAE* were used to assess the regression accuracy of each algorithm on each increased sample number dataset. Each dataset was evaluated five times with different random seeds (random_state= 1, 3, 9, 23, and 100), and each time the data for training and testing was split differently. The increased sample number dataset which showed the converged R^2, *RMSE,* and *MAE* values were selected as the appropriate increased sample number.

7) Model development

After the appropriate fluctuation range and increased sample number were settled, the final dataset to develop the ML models was constructed. The learning model was developed and run in Jupyter Notebook (version 6.0.1) on a laptop (Surface Laptop 2: Core i5-8250U CPU and 8 GB RAM, Microsoft) as described in section 2-4). RF, ET, GBDT, and XGBoost algorithms were implemented using the scikit-learn package (version 0.24.1) in Python (version 3.7.4). The dataset was imported and randomly split into two groups:

80% of the data was used for training the model and 20% was used for testing. The variable "random_state" was set to a fixed value to ensure the same split of the data for each model. The function "GridSearchCV" was applied to search for the optimal combination of hyperparameters for each model. The R^2, *RMSE*, and *MAE* values were calculated to assess the regression accuracy of each model, and the values were compared. Additionally, to further assess the performance of the models, the relative error [19] was defined as:

$$Relative\ error = \frac{|\hat{y}^{(i)} - y^{(i)}|}{y^{(i)}}, \quad (4)$$

where $y^{(i)}$ is the flexural strength calculated from *in vitro* experiments, $\hat{y}^{(i)}$ is the predicted flexural strength from the above four models.

8) ML algorithms basis

The selected algorithms are ensemble algorithms. In an ensemble, a set of base learners are trained to act together as a strong learner, thereby providing more accurate predictions [26]. Bagging and boosting are the two most frequently used approaches for constructing ensemble algorithms. The schematic stucture of bagging and boosting are shown in Figure 2. The ensemble algorithms used in this study included RF, ET, GBDT, and XGBoost. RF and ET algorithms use the bagging approach, and the rest of the two algorithms use boosting approach. A brief introduction of the four algorithms implemented in this study is provided below.

Random Forest (RF)

Breiman [37] developed the RF algorithm for both regression and classification purposes. As the base constituents of the ensemble are tree-structured predictors, and since each tree is constructed using an injection of randomness, the method is called "random forests" [42]. This tree-structured predictor is called a decision tree (DT). DT works by continuously splitting data into smaller and smaller subsets by calculating a certain parameter. The final subset of a regression tree gives a quantitative prediction, while the classification tree gives categorical predictions [43]. A bootstrap sample was used to train each decision tree as the basic estimator. The number of the decision tree and how the tree grows are controlled by the hyperparameters in the algorithm. Each decision tree will give a predicted value for the regression problem. The final output value in RF model is the unweighted average of all the predicted values obtained from all decision trees, which could be expressed as follows:

$$\hat{Y} = \frac{1}{t}\sum_{i=1}^{t} \hat{y}^{(i)} \quad , \tag{5}$$

where $\hat{Y}$ is the output value from RF model, t is the number of DT, and $\hat{y}$ is the prediction value of each decision tree.

Extra Tree (ET)

Geurts *et al*. [38] developed the ET algorithm as an extension from the RF algorithm.

ET is also a tree-based ensemble algorithm, which has some differences from that of RF. ET algorithm uses the whole training dataset to train each decision tree, while RF algorithm uses the bootstrap sample. RF algorithm obtains the best feature and value to split into two branches for decision tree by calculating specific mathematical parameter (*MSE* for regression problems), while ET algorithm obtains the splitting feature and value completely randomly.

Gradient Boosting Decision Tree (GBDT)

GBDT is a widely used ML algorithm proposed by Friedman [39] which integrates multiple DTs into a strong final ensemble model using the boosting approach [44, 45]. The training set was adaptively updated according to the performance of a previously created DT (predecessor). Each basic model is created sequentially trying to correct its predecessor. The process could be expressed as follows:

$$f_t(x) = f_{t-1}(x) + \alpha h_t(x) \ , \tag{6}$$

where $f_t(x)$ is the GBDT model and $h_t(x)$ is the basic model (DT) at step t , x is the configuration that need to be optimized ("chemical composition" in this study) and α is called the learning rate, which is a regularization parameter [26]. It scales the length of the step for finding the optimum solution. Larger alpha leads to faster iteration speed, while smaller alpha leads to lower iteration speed, which is more possible to find the optimum

solution, but it requires more computational cost. Except from regularization through shrinkage of the contributed basic models, randomness is also incorporated as an integral part of the fitting procedure.

eXtreme Gradient Boosting (XGBoost)

Under the framework of GBDT, XGBoost has been proposed with higher computation efficiency and better capability to deal with overfitting problems [40]. There are some main differences between GBDT and XGBoost, one is the different definition in objective function, in which a regularization function is implemented alongside the loss function.

$$Obj^{(t)} = \sum_{k=1}^{n} l(\bar{y}_i, y_i) + \sum_{k=1}^{t} \Omega(f_i) , \quad (7)$$

where l is the loss function, n is the number of observations used, t is step and Ω is the regularization term to prevent overfitting issue [46]. The other main difference from XGBoost is that GBDT only uses the first-order derivative information of the loss function when optimizing the objective function, while XGBoost performs a second-order Taylor expansion on the loss function, and both the first-order and second-order derivatives are used [26, 44].

3. Results

1) Appropriate fluctuation range

The R^2, *RMSE*, and *MAE* values of different fluctuation range datasets for each model are shown in Table 4 to Table 7**.** For the four models, fluctuation range 0.1 showed the highest R^2 value and the lowest *RMSE* and *MAE* value (RF: R^2: 0.972±0.015, *RMSE*: 5.199±1.277, *MAE*: 4.048±1.043; ET: R^2: 0.995± 0.004, *RMSE*: 2.172±0.900, *MAE*: 1.533±0.710; GBDT: R^2: 0.999±0.001, *RMSE*: 0.678±0.374, *MAE*: 0.435±0.190; XGBoost: R^2: 0.956±0.029, *RMSE*: 6.070±1.919, *MAE*: 3.805±1.067). Therefore, 0.1 was decided to be the appropriate fluctuation range. For GBDT and RF models, the R^2 value dropped below 0.9 starting from a fluctuation range of 0.8. For ET and XGBoost models, the R^2 value dropped below 0.9 starting from a fluctuation range of 0.7.

2) Appropriate increased sample number

For ET, GBDT, and XGBoost models, the R^2, *RMSE*, and *MAE* values showed convergence when the sample number increased to eight, while for the RF model, the evaluation metrics started to converge when the sample number increased to nine. The standard deviation value showed a decreased tendency from 1 sample to 11 samples (Figure 3 to Figure 6).

3) Tuned hyperparameters

The tuned hyperparameters for the final developed four models are shown in Table 8. For RF model, the hyperpameters “n_estimators”, “max_depth”, “min_samples_leaf”, “min_samples_split”, “max_features”, “min_weight_fraction_leaf” and “random_state” were adjusted. For ET model, “n_estimators”, “max_depth”, “min_samples_leaf”,

"min_samples_split", and "random_state" were adjusted. For GBDT model, "n_estimators", "max_depth", "learning_rate", "min_samples_leaf", "min_samples_split", "min_weight_fraction_leaf" and "random_state" were adjusted. For XGBoost model, "n_estimators", "max_depth", "learning_rate", "booster", "min_child_weight", "reg_alpha", "colsample_bytree", "colsample_bylevel" and "random_state" were adjusted.

4) Prediction performance of the four ML models developed

The prediction performance of each model is shown in Table 9. The R^2 values for RF, ET, GBDT, and XGBoost were 0.947, 0.997, 0.998, and 0.927, respectively. The relative errors of all the models were within 15%.

The flexural strength predictions using the ML models beyond the training set are plotted versus the observed values in Figure 7. The data points within the region of two black dashed lines have relative errors less than 5%. For ET and GBDT models, the test points were all within the black dashed lines. However, for RF and XGBoost models, two test points deviated from the black dashed lines.

4. Summary

All of the four models developed achieved R^2 values of over 0.9 in terms of prediction accuracy. GBDT was the best performing model, closely followed by the ET model.

III. Feature Importance

1. Purpose

After the ML models were developed, this chapter aims at conducting a feature importance analysis using the developed models to investigate the components that affect the flexural strengths of CAD/CAM RCBs.

2. Materials and methods

The feature importance analysis was conducted using scikit-learn package (version 0.24.1) in Python (version 3.7.4) on a laptop (Surface Laptop 2: Core i5-8250U CPU and 8 GB RAM, Microsoft). Feature importance analysis is a built-in attribute inside each of the four algorithms. It determined the contribution of each feature to the targeted properties (*i.e.*, flexural strength) by assigning a score for each feature. The importance of a feature is basically: how much this feature is used in each tree of the forest. It was calculated by implementing the function "feature_importances_", thereafter, the scores were calculated automatically. Finally, the scores were compared for each component within the ET and GBDT models.

3. Results

The feature importance analysis result is shown in Figure 8. The filler content, TEGDMA, and $ZrSiO_4$ were the top three features that had a relatively high importance value in ET and GBDT models. However, even though SiO_2 was contained in all products, it had little importance for ET and GBDT models. By contrast, as a synthetic amorphous

form of silicon dioxide, micro-fumed silica had relatively high importance, ranking fourth for the ET model and fifth for the GBDT model.

4. Summary

The feature importance analysis demonstrated that the filler content, TEGDMA, and $ZrSiO_4$ were the top three important features.

IV. Exhaustive search

1. Purpose

The feature importance analysis conducted in Chapter III revealed which feature had great influence on flexural strength. However, influence of each feature is positive or negative has not been clarified. Therefore, an exhaustive search was conducted in this chapter to illustrate the above mentioned influence of the component and to identify effective compositions.

2. Materials and methods

1) Running environment

The exhaustive search was conducted in Jupyter Notebook (version 6.0.1) on a laptop (Surface Laptop 2: Core i5-8250U CPU and 8 GB RAM, Microsoft) using scikit-learn package (version 0.24.1) in Python (version 3.7.4).

2) Exhaustive search

The first 14 descriptors were set to 0 or 1, and the 15th descriptor (filler content) was set to be in the range of 62 to 85. The function "itertools.product" was used to generate $2^{14} \times 24$ (393,216) combinations of the descriptors. An exhaustive search was performed by GBDT model to iterate all $2^{14} \times 24$ (393,216) combinations of the descriptors and predict 393,216 values of the corresponding flexural strengths. By checking the original output of the exhaustive search, the predicted flexural strength and its combination number were plotted. The composition differences were analyzed among the high and low predicted

flexural strengths groups, by checking how many times a specific component was contained among those compositions. Accordingly, the frequency for each component is defined as:

$$Frequency = \frac{Times\ a\ specific\ component\ occurs}{Number\ of\ all\ predicted\ compostions} \quad (8)$$

Finally, the possible compositions that lead to the highest predicted flexural strength were illustrated.

3. Results

1) Original output of the exhaustive search

Part of the original output of the exhaustive search is shown in Figure 9. The original output of the exhaustive search contains three parts: the combination numbers, the corresponding compositions, and the predicted flexural strengths. The combination numbers are from 1 to 393,216, and some of these numbers are displayed as examples in the red square. For each combination, it represents a specific composition for the CAD/CAM RCB that was formed by 0 and 1 as shown in the yellow square. By checking the corresponding feature name, we could know what is contained in this composition and its filler content. In the meantime, the model will predict the flexural strength for each combination, which was shown in the blue square.

2) Plot of the exhaustive search

Figure 10 shows the plot of an exhaustive search using the GBDT model. The horizontal

axis of Figure 10 represents the combination number (red square of Figure 9), and the vertical axis represents the corresponding flexural strength predicted by the model (blue square of Figure 9). The plot of the GBDT model mainly contains three groups of predictions, with the top group and the bottom group indicated by red and red dotted circles. For the GBDT model, the bottom group of Figure 10 shows four high-density areas of points, where the predictions varied from 132.2 MPa to 180.1 MPa (93,135 compositions). By contrast, the highest predictions in the top group ranged from 256.6 MPa to 269.5 MPa (4,527 compositions).

3) Composition differences between higher and lower predictions group

The composition differences in the higher prediction group and lower prediction group are demonstrated in Figure 11. Among the compositions in higher prediction group, the filler content ranged from 81 to 85 (wt)%, and ranged from 62 to 69 (wt)% in the lower prediction group. For the monomers in the lower predictions group, TEGDMA shared the highest ratio of 53.5%. In the higher predictions group for GBDT model, none of the combinations contained TEGDMA or Bis-GMA; instead, UDMA and Bis-EMA were the most frequently contained monomers. In the case of fillers, $ZrSiO_4$ and micro-fumed silica had a relatively high frequency in the lower prediction group for GBDT model predictions (59.2% and 60.6%, respectively), whereas it dramatically decreased in the higher prediction group.

4) The maximum prediction by GBDT model and its corresponding composition

The maximum flexural strength predicted by the GBDT model was 269.5 MPa. For the GBDT model's prediction, the related chemical composition for the maximum flexural strength was SiO_2, barium glass, methacrylate mixed filler, with or without Al_2O_3 for fillers, and UDMA alone for monomers. Simultaneously, the filler content was from 82 to 85 (wt)%.

4. Summary

The exhaustive search showed UDMA and Bis-EMA were the most frequently contained monomers, and the filler content ranged from 82 to 85 (wt)% in the higher prediction group. Effective compositions predicted by GBDT model were identified.

V. Bayesian optimization

1. Purpose

In Chapter IV, an exhaustive search was applied and effective compositions were identified. However, the exhaustive search is a time-consuming process, and therefore, a Bayesian optimizer was used to shorten the time to search for the effective composition.

2. Materials and methods

1) Building composition space

To implement Bayesian optimization for chemical reaction parameter selection, we used a user-friendly package "Experimental Design via Bayesian Optimization" (short for EDBO) developed by Shields *et al*. [47] for our data. First, a dictionary of components was defined following the same rule as exhaustive search, the first 14 components were set to 0 or 1, and the 15th descriptor (filler content) was set to be in the range from 62 to 85 (wt)%. Then, a dictionary of desired encodings was specified; the first 14 components used one-hot encoding (OHE) [47]. This method details the presence or absence of certain components with 0 or 1; the 15th descriptor was numeric. As the next step, a Bayesian optimizer, called "BO_express" inside this package was applied and instantiated to automatically build the composition space.

2) Surrogate model and acquisition functions

Over the course of our study, we investigated three acquisition functions, including mean maximization (MeanMax), variance maximization (VarMax), and expected improvement (EI). MeanMax is a purely exploratory algorithm which selects compositions for

optimization via domain points with the highest model variance. MaxVar is a purely exploratory algorithm which selects compositions for optimization via domain points with the highest model variance. EI attempts to balance exploration and exploitation by accounting for the amount of improvement over the best-observed value. The surrogate model was Gaussian process regression, which was already built in the "BO_express". "BO_express" has built-in adjusted hyperparameters for different encoding methods. We select the hyperparameters for OHE method.

3) Initialization

From the exhaustive search space, five random compositions (the batch size was chosen to be five according to [47]) and their corresponding flexural strengths predicted by GBDT model were used as initialization data. After the surrogate model was updated according to these five predictions, the acquisition function was applied to propose five new candidate compositions. These newly proposed compositions were then given to GBDT model to make predictions. Afterwards, these five compositions and their predicted flexural strengths were added to the observations, and the surrogate model updated again, then another five new candidate compositions were proposed by the acquisition function. This process was repeated 10 times, calling each time as one "iteration". Three different initializations were performed. The time to find the highest flexural strength among combinations was compared between the exhaustive search optimized by BO and without BO.

4) Bayesian optimization basis

Bayesian optimization (BO) is an approach to optimizing objective functions that take a long time. Instead of searching all the possible combinations like exhaustive search, the BO method could propose candidate compositions that have a high possibility to lead to improved flexural strength, therefore, reducing the predicting time for GBDT model. It builds a surrogate for the objective and quantifies the uncertainty in that surrogate using a Bayesian machine learning technique, and then uses an acquisition function defined from this surrogate to decide where to sample [48]. Sequential model-based Bayesian optimization (SMBO) is one of the Bayesian optimization approaches [49]. The framework is as follows [50]:

SMBO (f, M_0, T, S)

$H \leftarrow \varnothing$,

For $t \leftarrow 1$ **to** T,

  $x^* \leftarrow \text{argminx } S(x, M_{t-1})$,

  Evaluate $f(x^*)$,

  $H \leftarrow H \cup (x^*, f(x^*))$,

  Fit a new model M_t to H.

return H

f is the target black box function that needs to be optimized, usually, it is very time-consuming to calculate f. Therefore, a surrogate S is usually used to "mimic" f, which requires less calculation compared to f. x are the combinations of the hyperparameters. H is the observed history. To pick the next evaluated hyperparameter x^*, S needs to be minimized, to minimize S, usually an acquisition function should be applied and

maximized. *M* is the fitting distribution model.

The commonly used probability distribution model (surrogate model) is Gaussian process regression, which estimates the mean and variance of the training data as a posterior distribution [48, 51]. However, the dimensions of the descriptors could be large, resulting in a large reaction space of up to tens of thousands of possible compositions that cannot be all conducted *in vitro* to update the posterior distribution. Therefore, after training the surrogate model, an acquisition function is used to select the next trial experiment from the reaction space [52]. There are two typical strategies in acquisition functions: exploration and exploitation. Exploitation tends to select the next experiment around the neighborhood of the current best-observed value, while exploration tends to select the next point with the greatest predictive uncertainty, and tends to investigate the entire reaction space thoroughly [47]. As in any iterative search algorithm, the goal is to balance exploration of options for *x* with exploitation of previously-explored options, so that a good choice is found in a small number of trials [49].

Commonly used acquisition functions such as expected improvement (EI) aim to balance these two strategies. EI was proposed initially in Mockus *et al*. (1978) [53] and then made popular by Jones *et al*. (1998) [54]. EI measures the expectation of the improvement on *f* (*x*) with respect to the predictive distribution of the probabilistic surrogate model. EI considers not only the probability of improving the current best estimate but also factors in the magnitude of the improvement. The point with the highest expected improvement will be selected as the next candidate. EI has the form [55] :

$$EI(x) = \begin{cases} (f(x^+) - \mu(x) - \xi)\Phi(Z) + \sigma(x)\varphi(Z) if\ \sigma(x) > 0) \\ 0\ if\ \sigma(x) = 0 \end{cases} \quad (9)$$

And

$$Z = \begin{cases} \dfrac{f(x^{+}) - \mu(x) - \xi}{\sigma(x)} \; if \; \sigma(x) > 0 \\ 0 \; if \; \sigma(x) = 0 \end{cases} \tag{10}$$

where φ(Z) and Φ (Z) are the probability distribution and the cumulative distribution of the standardized normal, respectively. $f(x^{+})$ is the best value of the objective function observed so far, μ(x) and σ(x) are the mean and standard deviation of the probabilistic surrogate model. ξ is an empirical parameter that can be actively managed for the trade-off between exploration and exploitation. the next point to evaluate is chosen according to: $x_{n+1} = argmax_{x \in X} EI(x)$. A graphical overview of Bayesian optimization is shown in Figure 12.

3. Results

1) Performance of different acquisition functions

With three times different initializations, EI all demonstrated the best performance (Figure 13). For each initialization, EI found the highest flexural strength (269.5MPa) among combinations as concluded from exhaustive search within 10 iterations. MeanMax and VarMax did not achieve the desired performance.

2) Time reduction by EI

As shown in Figure 14, in the first initialization, EI found the highest flexural strength

in the 9th iteration (6.00 hours). In the second and third initializations, EI found the highest flexural strength in the 10th and 6th iterations, respectively (6.67 hours and 4.00 hours, respectively). The average time to find the 269.5 MPa and its corresponding composition is 5.56 hours, which is decreased by 8 times compared to the time for exhaustive search (48 hours).

3) Comparison of the compositions proposed by EI and exhaustive search

The detailed composition proposed by BO in the 9th iteration of the first initialization is shown in Figure 15. In figure 16, the compositions proposed by BO of 10 iterations in the first initialization were plotted on the exhaustive search figure. It can be seen that the proposed compositions from BO encompass a varied range of predicted flexural strengths. Furthermore, EI predominantly focused its searches on the top prediction group of exhaustive search, as indicated by the higher density of orange compared to the other areas in Figure 16. Finally, the composition proposed by BO that led to the 269.5 MPa (Figure 15) was shown in the white dot in Figure 16.

4. Summary

BO with EI as an acquisition function shortened the time (5.56 hours) for GBDT model to find an effective composition compared to exhaustive search (48 hours) regardless of the initialization data.

VI. Discussion

AI technology has been applied to the dental field mainly for the diagnosis, decision-making, and treatment planning [56-60]. Farhadian *et al*. [58] developed a support vector machine model and diagnosed three classifications of periodontal diseases using data collected from 300 patients. Reyes *et al*. [59] developed caries prognosis models in primary and permanent teeth after 2 and 10 years of follow-up through a ML approach, using predictors collected in early childhood. Sakai *et al*. [60] developed a deep neural network based on LeNet-5 and predicted the implant treatment plans using cone beam computed tomography images collected from 60 patients. However, it has not yet been used to modify the properties of dental materials. In this study, CAD/CAM RCBs was chosen as the target material for the application of the interpretable AI process. It is an ideal material for AI modelling study, since the mechanical properties of CAD/CAM RCBs could be improved simply by modification of the compositions. This study was performed to establish a process using AI technology to predict the flexural strength of CAD/CAM RCBs and investigate the components that affect flexural strength to explore the optimum composition based on the available dataset.

Data size is of great importance in the ML process [28]. It is likely that relatively little attention has been paid to the application of ML methods to dental materials, particularly CAD/CAM RCBs, because of the limitation of compositional information. Therefore, before developing the final ML models, we conducted the fluctuation range and sample increment study. For GBDT and RF models, the R^2 value dropped below 0.9 starting from a fluctuation range of 0.8. For ET and XGBoost models, the R^2 value dropped below 0.9 starting from the fluctuation range of 0.7. This result could be explained by the overlap of

the fluctuated range for both 0 and 1. Specifically, when the fluctuation range was 0.7, 1 could be represented with random numbers within the range of [0.65, 1.35], while 0 could be represented within the range [0, 0.7]. The overlapping of the ranges may result in the models erroneously interpreting the inclusion or exclusion of the related component, thereby reducing the accuracy of predictions.

In the sample increment study, increasing only one or two samples resulted in minus R^2 value and large standard deviation, which indicated the unstable performance of the models. ET, GBDT, and XGBoost models showed converged evaluated metrics on increased sample number of eight, but RF showed the convergence on increased sample number of nine. When considering nine newly generated samples, the total sample count would reach 120, making it more favorable for subsequent processes such as the 10-fold cross-validation due to its divisibility. Based on the above reasons, the increased sample number was determined to be nine.

Furthermore, to avoid the negative effect of a small dataset, ensembles were used in this study. In an ensemble, a set of base learners are trained to act together as a strong learner, thereby providing more accurate predictions [26]. Bagging and boosting are the two most frequently used approaches for constructing ensemble models [61]. Among these two approaches, algorithms that involve bagging, such as RF and ET, and algorithms that use boosting, such as GBDT and XGBoost, have demonstrated robust performance in predicting material properties in recent studies [62-64]. One study demonstrated R^2 values of over 0.939 in predicting the compressive strength of concrete using ensemble algorithms, utilizing a dataset comprising 154 data points [26]. In the present study, the four selected ensemble algorithms all achieved acceptable R^2 values of over 0.9 in terms of prediction

accuracy. The *RMSE* and *MAE* values indicated the superior predictive ability of the developed models for the flexural strength of CAD/CAM RCBs. GBDT was the best-performing model. This was followed closely by the ET model. These results imply that, for the available dataset applied in this study, the algorithms performed similarly when bagging or boosting techniques were used.

Because the ML models were only trained on limited data, it was crucial for the models to provide reliable predictions beyond the training set. The prediction results using the test data suggested that the ET and GBDT models had a promising ability to reliably predict the flexural strength values of CAD/CAM RCBs, provided the chemical compositions were given. However, several predictions of the XGBoost and RF models deviated from the observed values, and their relative errors were within 15%. Moreover, the *RMSE* values for the RF and XGBoost models were 5.64 MPa and 6.64 MPa, respectively, which were much larger than those for the other two models. These results suggest that the RF and XGBoost models yielded uncertainties during the extrapolation process.

The translation of components into a machine-readable format can play a critical role in the development of predictive models [65]. In Chapter I, OHE method was used to represent each component. This computationally simplistic method merely details the presence or absence of certain components with 0 or 1, whilst encoding no chemical information. Some studies implemented other representation methods for the inclusion of structural and chemical information for the components; for instance, Morgan fingerprints [66] (atom types, neighbouring connectivity…), Morderd descriptors [67, 68] (atom-bond connectivity index, acidic group count…) and Density Functional Theory (DFT) Descriptors [69, 70] (molar mass, electronic spatial extent…). Pomberger *et al.* [66]

demonstrated a decrease of 2% in *RMSE* value of RF model when using Morgan fingerprints to generate descriptors for the input composition to predict the yield of their target reaction compared to OHE, while no improvement for *RMSE* when using all ligand DFT descriptors. The utilization of the structural and chemical information representation method can lead to an increase in the number of descriptors, subsequently necessitating a larger sample size. Given the limited dataset available, the simplest OHE method was used in this study. This decision was made by the need to accommodate the data constraints while still enabling meaningful interpretation for the next analysis on exhaustive search.

After the ML models were developed, in Chapter III and IV, the effective components that contribute to the flexural strength of CAD/CAM RCBs was investigated. In terms of the feature importance analysis, the high importance of the top three features were filler content, TEGDMA, and $ZrSiO_4$. According to the ranking, the filler content and the use of monomers should be adjusted as a priority when modifying the composition of CAD/CAM RCBs to improve flexural strength. Furthermore, even though SiO_2 is contained in all products, it had little importance for all models. By contrast, as a synthetic amorphous form of silicon dioxide, micro-fumed silica had relatively high importance, ranking fourth for ET model and fifth for the GBDT model, which suggests that the different size and manufacturing process for the same element contributed differently to the flexural strength of CAD/CAM RCBs. However, the feature importance could not illustrate whether these components demonstrated positive or negative contribution to the flexural strength of CAD/CAM RCBs. Therefore, the exhaustive search was conducted to reveal the influence in detail in Chapter V.

The exhaustive search results of GBDT model were analyzed because of its promising

prediction performance. In the bottom group of predictions from the GBDT model, the filler content in the combinations ranged from 62 to 69 (wt)%, whereas in the top group, it ranged from 82 to 85 (wt)%, yielding higher predicted flexural strength values. It has been proved that high filler content improves flexural strength [71], which is in agreement with the finding in the present study. However, exceptions to this rule were observed for MZ 100, which obtained a relatively low strength value compared with other materials listed in this study, despite its relatively high filler fraction of 85 (wt)%. Furthermore, the same predictions were made by the ML models using compositions that varied only in filler content. This result suggests that, for CAD/CAM RCBs, the filler content and flexural strengths did not always positively correlate. Flexural strength tended to decrease when the filler content exceeded 60 (vol)% [5, 72]. Our results suggest that, within a specific range, increasing/decreasing the content of fillers does not necessarily improve/decrease the flexural strength.

For the monomers in the lower predictions group, TEGDMA shared the highest ratio of 53.5%, This result suggests that TEGDMA could be an essential component that affects the predicted flexural strength. This finding is in agreement with that obtained by *in vitro* results [73], where increasing the content of TEGDMA in the experimental RCBs slightly reduced the flexural strength.

In the higher predictions group for GBDT model, none of the combinations contained TEGDMA or Bis-GMA; instead, UDMA and Bis-EMA were the most frequently contained monomers. UDMA has a lower molecular weight and exhibits lower viscosity than Bis-GMA, and simultaneously demonstrates higher cross linking and polymerization activity with light curing [5, 74]. UDMA can be used alone or with other monomers, such

as TEGDMA and Bis-GMA, in a matrix. For an *in vitro* experiment, it was reported that when fillers were fixed, UDMA used alone exhibited the highest flexural strength compared with the mixture of different ratios of UDMA and TEGDMA [72]. The present result was consistent with this finding. Bis-EMA is a monomer analogous to Bis-GMA, but without the two hydroxyl groups (-OH). It is more flexible and mobile than Bis-GMA [75]. Therefore, it demonstrates a higher overall conversion [76-78]. Another study has reported that UDMA/Bis-EMA/TEGDMA monomer mixtures are characterized with better flexural strength compared with UDMA/TEGDMA [78]. However, in this study, the model failed to identify the underlying collective pattern within these three components. This discrepancy could potentially be attributed to the absence of specific monomer combinations (as mentioned above) in the original input products.

Furthermore, the results of the exhaustive search also indicated that among all the combinations containing Bis-EMA, half also included UDMA. These findings suggest that Bis-EMA alone may not contribute to the heightened flexural strength; rather, the enhanced strength might be a result of the synergistic effects of its combination with UDMA. This observation aligns with the feature importance analysis, where Bis-EMA does not exhibit a notably high importance score.

In the case of fillers, the results demonstrated that $ZrSiO_4$ had a relatively high frequency in the lower prediction group for GBDT models' predictions (59.2%), whereas it halved in the higher prediction group. $ZrSiO_4$ is an important component in the ceramic industry because of its high chemical stability and excellent coloring performance at high firing temperatures [79]. These results suggest that $ZrSiO_4$ may have some effect on the flexural strength of CAD/CAM RCBs, which needs to be explored.

It is noteworthy that the maximum flexural strength predicted by the GBDT models was 269.5 MPa. For the GBDT model's prediction, the related chemical compositions for the maximum flexural strength were SiO_2, barium glass, methacrylate mixed filler, with or without Al_2O_3 for fillers, and UDMA alone for monomers. Simultaneously, the filler content ranged from 82 to 85 (wt)%. None of these combinations were shown in Table 1; therefore, these compositions could be considered for use in the future.

BO has become prominent especially as a tool to optimize hyperparameters of ML models and has become of interest to the chemical community [80-83]. In Chapter V, a Bayesian optimizer and its user-friendly platform (EDBO) were applied to shorten the time for exhaustive search to find an effective composition which led to the highest predicted flexural strength (269.5 MPa). This optimizer was originally developed for optimizing chemical reaction configurations. Shields *et al.* [47] used the BO method to optimize the yield of two reactions in the pharmaceutical field and successfully found unconventional compositions and configurations that were not commonly selected by human experts and improved the reaction yield within only 40 experiments. Rankovi´c *et al*. [84] also reported that when BO was applied on a 720 additives dataset, the effective additives were proposed as evidenced by the high-throughput experimentation in less than 100 iterations. In this study, within three different initializations, the Bayesian optimizer proposed the effective composition within only 30 composition predictions in minimum (Maximum: 50). Compared with searching 393,216 compositions in exhaustive search (48 hours), implementing this Bayesian optimizer decreased the time to an average of 5.56 hours. These results agree with former studies on the effectiveness of the BO method in reducing the search time for effective compositions.

During the process of three different initializations, only the acquisition function EI successfully proposed the effective composition within 10 iterations, which suggested that balance exploration and exploitation is an effective way in this study. The robust performance of EI were also demonstrated by other studies in proposing effective molecules, reaction parameters, and structures for different materials [85-87]. Furthermore, in this study, the results of MeanMax in the first and second optimization processes did not show any improvement even after the second and fifth iterations. This result could be explained by the fact that MeanMax (pure exploitation) tends to select the next point near the temporary best value, making it easy to be trapped in local maxima. By contrast, VarMax (pure exploration) tends to thoroughly investigate the entire composition space. However, it may not necessarily find the global minima without searching the entire space, which can be very time-consuming. In the second and third optimization process, VarMax's proposed composition close to the 269.5 MPa. Therefore, it is assumed that VarMax will finally be converged if iterations are added.

Within the scope of this study, Chapters IV and V presented two approaches for proposing effective compositions of CAD/CAM RCBs. While BO offers a time-efficient means of identifying such compositions, an exhaustive search provides a wealth of information that can be analyzed. Specifically, it elucidates the influence of each component on the flexural strength of CAD/CAM RCBs by examining and comparing the compositional differences of compositions in the high and low predicted flexural strengths groups. Hence, an exhaustive search can be recommended under the scenario that deals with a small compositional space and enables calculations to be completed within a few days, and also when the influence of each descriptor on the target property is needed.

However, if the goal is only to discover effective formulations, it is advisable to employ the BO method.

Regarding the limitation of this study, the data used for training the ML models in Chapter II were not sufficiently comprehensive to contain all factors that affected the target material property. For instance, the detailed concentrations of each component, the temperature and pressure used during the manufacturing process, the size and shape of the fillers. Therefore, these factors should be considered for further improvement. Even though an approach was established to increase the dataset, the information contained was still limited; hence, the models developed in this study may not have excellent extrapolating ability. Therefore, the newly released products and their compositional information should be further added. In Chapter V, the hyperparameters for the surrogate model were chosen directly from the literature, as they were optimized specifically for the OHE representation and demonstrated satisfactory results in our study. However, if the hyperparameters were adjusted using the dataset developed in Chapter I prior to their application, it would further enhance the comprehensiveness of the interpretable AI process developed in this study. Furthermore, the ML model developed in this study may be applicable for other properties of CAD/CAM RCBs, such as hardness and water absorption ability, and they should also be considered in the future for enhanced clinical application.

VII. Conclusion

In this study, the interpretable AI process was established, and flexural strength of CAD/CAM RCBs were predicted with acceptable accuracy. UDMA and filler content were identified by the process to be the major factors to affect the flexural strength during the established process. This technology has great potential as an important tool to help modify various targeting properties for different dental materials, provided an adequate data set is available, which will save time and cost for materials designing.

VIII. Acknowledgement

During my Ph.D. journey spanning four years, I find myself profoundly grateful for the numerous individuals who have generously extended their support, encouragement, and friendship, thus transforming my time in Japan and in Osaka into an unforgettable and cherished chapter of my life. Their unwavering presence and guidance have left an indelible mark on my heart, filling it with gratitude and profound appreciation.

I would like to express my sincere appreciation to Professor Satoshi Imazato, Chair of the Department of Dental Biomaterials at Osaka University Graduate School of Dentistry, for accepting me as his student. Without his mentorship, I would not have embarked on this fulfilling journey of research and ignited my true passion. I am deeply grateful for Professor Imazato's kind, patient, and dedicated guidance throughout my academic pursuit. His wisdom and warmth shine through our discussions, as I consistently gain new knowledge and receive valuable advice. His influence extends beyond research, as he imparts valuable life lessons, emphasizing the importance of curiosity, critical thinking, and nurturing lifelong hobbies. I sincerely thank him for being an inspiring role model who has made a profound impact on my academic and personal growth.

I would like to express my heartfelt gratitude to Associate Professor Satoshi Yamaguchi for providing me with detailed and valuable guidance on my research. His immense knowledge opened a brand new and expansive field for me, broadening my horizons. And a lot of thanks to his gentle support and understanding of my emotion when I got frustrated and lost motivation. I really appreciate him for being always positive towards both me and

the research, his attitude affects me and supports me to go through the tough days of research work. Thanks to both him and Imazato sensei, I built confidence for myself and towards research. I will keep challenging myself.

To my parents and grandparents, thanks for them being understanding and supporting me in my decision to pursue a Ph.D. Their love and care are indispensable.

To my husband, thanks for his love and support. Whenever I was down, he is always there and stand by my side, providing me with comfort and strength. I am deeply grateful for his understanding and encouragement in pursuing my goal of becoming a researcher.

Finally, I would say, my graduation and this Ph.D. dissertation are not the end, but the beginning.

IX. References

[1] Ruse ND, Sadoun MJ. Resin-composite blocks for dental CAD/CAM applications. J Dent Res 2014;93(12):1232-4.

[2] Miura S, Kasahara S, Yamauchi S, Katsuda Y, Harada A, Aida J, Egusa H. A possible risk of CAD/CAM-produced composite resin premolar crowns on a removable partial denture abutment tooth: a 3-year retrospective cohort study. J Prosthodont Res 2019;63(1):78-84.

[3] Kabetani, T, Ban S, Mine A, Ishihara T, Nakatani H, Yumitate M, Yamanaka A, Ishida M, Matsumoto M, Meerbeek BV, Shintani A, Yatani H. Four-year clinical evaluation of CAD/CAM indirect resin composite premolar crowns using 3D digital data: Discovering the causes of debonding. J Prosthodont Res 2022;66(3):402-408.

[4] Lee C, Kashima K, Ichikawa A, Yamaguchi S, Imazato S. Influence of hydrolysis degradation of silane coupling agents on mechanical performance of CAD/CAM resin composites: In silico multi-scale analysis. Dent Mater J 2020;39(5):803-7.

[5] Mainjot AK, Dupont NM, Oudkerk JC, Dewael TY, Sadoun MJ. From artisanal to CAD-CAM Blocks: State of the art of indirect composites. J Dent Res 2016;95(5):487-95.

[6] Tsujimoto A, Barkmeier WW, Takamizawa T, Latta MA, Miyazaki M. Influence of thermal cycling on flexural properties and simulated wear of computer-aided design/computer-aided manufacturing resin composites. Oper Dent 2017;42(1):101-10.

[7] Liu Y, Zhao TL, Ju WW, Shi SQ. Materials discovery and design using machine learning. J Materiomics 2017;3(3):159-77.

[8] Shan T, Tay FR, Gu L. Application of artificial intelligence in dentistry. J Dent Res 2021;100(3):232-44.

[9] Yamaguchi S, Lee C, Karaer O, Ban S, Mine A, Imazato S. Predicting the debonding of CAD/CAM composite resin crowns with AI. J Dent Res 2019;98(11):1234-8.

[10] Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: Chances and challenges. J Dent Res 2020;99(7):769-74.

[11] Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature 2018;559(7715):5[47]-55.

[12] Chen CT, Gu GX. Machine learning for composite materials. Mrs Commun 2019;9(2):556-66.

[13] Bzdok D, Altman N, Krzywinski M. POINTS OF SIGNIFICANCE Statistics versus machine learning. Nat Methods 2018;15(4):232-3.

[14] Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang GZ. XAI-Explainable artificial intelligence. Sci Robot 2019;4(37):eaay7120.

[15] Arrieta AB, Diaz-Rodriguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inform Fusion 2020;58:82-115.

[16] Molnar C, Casalicchio G, Bischl B. Interpretable Machine Learning - A Brief History, State-of-the-Art and Challenges. Comm Com Inf Sc 2020;1323:417-31.

[17] Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B. Definitions, methods, and applications in interpretable machine learning. P Natl Acad Sci USA 2019;116(44):22071-80.

[18] Ward L, Agrawal A, Choudhary A, Wolverton C. A general-purpose machine learning framework for predicting properties of inorganic materials. NPJ Comput Mater 2016;2(1):1-7.

[19] Hu YJ, Zhao G, Zhang MF, Bin B, Del Rose T, Zhao Q, Zu Q, Chen Y, Sun XK, de Jong M, Qi L. Predicting densities and elastic moduli of SiO_2-based glasses by machine learning. NPJ Comput Mater 2020;6(1):25.

[20] Tso WW, Burnak B, Pistikopoulos EN. HY-POP: Hyperparameter optimization of machine learning models through parametric programming. Comput Chem Eng 2020;139:106902.

[21] Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing 2020;415:295-316.

[22] Hung M, Voss MW, Rosales MN, Li W, Su WC, Xu J, Bounsanga J, Ruiz-Negrón B, Lauren E, Licari WF. Application of machine learning for diagnostic prediction of root caries. Gerodontology 2019;36(4):395-404.

[23] Yan W, Renteria C, Huang Y, Arola DD. A machine learning approach to investigate the materials science of enamel aging. Dent Mater 2021;37(12):1761-71.

[24] Yamaguchi S, Li H, Imazato S. Materials informatics for developing new restorative dental materials: A narrative review. Front Dent Med 2023;4:1123976.

[25] Bishnoi S, Singh S, Ravinder R, Bauchy M, Gosvami NN, Kodamana H, Krishnan

Anoop NM. Predicting Young's modulus of oxide glasses with sparse datasets using machine learning. J Non-Cryst Solids 2019;524:119643.

[26] Marani A, Nehdi ML. Machine learning prediction of compressive strength for phase change materials integrated cementitious composites. Constr Build Mater 2020;265:120286.

[27] Xiong J, Shi SQ, Zhang TY. A machine-learning approach to predicting and understanding the properties of amorphous metallic alloys. Mater Design 2020;187:108378.

[28] Ramprasad R, Batra R, Pilania G, Mannodi-Kanakkithodi A, Kim C. Machine learning in materials informatics: recent applications and prospects. NPJ Comput Mater 2017;3(1):54.

[29] Ducke VM, Ilie N. Aging behavior of high-translucent CAD/CAM resin-based composite blocks. J Mech Behav Biomed Mater 2021;115:104269.

[30] Alamoush RA, Silikas N, Salim NA, Al-Nasrawi S, Satterthwaite JD. Effect of the composition of CAD/CAM composite blocks on mechanical properties. Biomed Res Int 2018;2018:4893143.

[31] Lauvahutanon S, Takahashi H, Shiozawa M, Iwasaki N, Asakawa Y, Oki M, Finger WJ, Arksornnukit M. Mechanical properties of composite resin blocks for CAD/CAM. Dent Mater J 2014;33(5):705-10.

[32] Zhi L, Bortolotto T, Krejci I. Comparative in vitro wear resistance of CAD/CAM composite resin and ceramic materials. J Prosthet Dent 2016;115(2):199-202.

[33] Abdou A, Takagaki T, Alghamdi A, Tichy A, Nikaido T, Tagami J. Bonding performance of dispersed filler resin composite CAD/CAM blocks with different

surface treatment protocols. Dent Mater J 2021;40(1):209-19.

[34] Niem T, Youssef N, Wostmann B. Influence of accelerated ageing on the physical properties of CAD/CAM restorative materials. Clin Oral Invest 2020;24(7):2415-25.

[35] ISO 6872:2015. Dentistry - Ceramic materials. International Organization for Standardization 2015.

[36] Lawson NC, Bansal R, Burgess JO. Wear, strength, modulus and hardness of CAD/CAM restorative materials. Dent Mater 2016;32(11):e275-e83.

[37] Breiman L. Random forests. Mach Learn 2001;45(1):5-32.

[38] Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn 2006;63(1):3-42.

[39] Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Stat 2001;29(5):1189-232.

[40] Chen TQ, Guestrin C. XGBoost: A scalable tree boosting system. Proc ACM SIGKDD Int Conf Knowl 2016:785-94.

[41] Gholamy A, Kreinovich V, Kosheleva O. Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation. Int J Intell Syst Technol Appl 2018;11(2):105-111.

[42] Segal MR. Machine learning benchmarks and random forest regression. eScholarship Repository. University of California.

[43] Tong WD, Hong HX, Fang H, Xie Q, Perkins R. Decision forest: Combining the predictions of multiple independent decision tree models. J Chem Inf Comp Sci 2003;43(2):525-31.

[44] Guelman L. Gradient boosting trees for auto insurance loss cost modeling and prediction. Expert Syst Appl 2012;39(3):3659-67.

[45] Friedman JH. Stochastic gradient boosting. Comput Stat Data An 2002;38(4):367-78.

[46] Fan JL, Wang XK, Wu LF, Zhou HM, Zhang FC, Yu X, Lu XH, Xiang YZ. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. Energ Convers Manage 2018;164:102-11.

[47] Shields BJ, Stevens J, Li J, Parasram M, Damani F, Alvarado JIM, Janey JM, Adams RP, Doyle AG. Bayesian reaction optimization as a tool for chemical synthesis. Nature 2021;590(7844):89-96.

[48] Frazier PI. A tutorial on Bayesian optimization. arXiv preprint arXiv:180702811 2018.

[49] Yogatama D, Smith NA. Bayesian optimization of text representations. arXiv preprint arXiv:150300693 2015.

[50] Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. Adv Neural Inf Process Syst 2011;24.

[51] Snoek J, Larochelle H, Adams RP. Practical bayesian optimization of machine learning algorithms. Adv Neural Inf Process Syst 2012;25.

[52] Shahriari B, Swersky K, Wang ZY, Adams RP, de Freitas N. Taking the human out of the loop: A review of Bayesian Optimization. P IEEE 2015;104(1):148-75.

[53] Mockus J. The application of Bayesian methods for seeking the extremum.

Towards global optimization 1998;2:117.

[54] Jones DR, Schonlau M, Welch WJ. Efficient global optimization of expensive black-box functions. J Global Optim 1998;13(4):455-92.

[55] Archetti F, Candelieri A. The acquisition function. In: Bayesian Optimization and Data Science, Cham: Springer International Publishing 2019, 57-72.

[56] Abdalla-Aslan R, Yeshua T, Kabla D, Leichter I, Nadler C. An artificial intelligence system using machine-learning for automatic detection and classification of dental restorations in panoramic radiography. Oral Surg Oral Med Oral Pathol Oral Radiol 2020;130(5):593-602.

[57] Alabi RO, Elmusrati M, Sawazaki-Calone I, Kowalski LP, Haglund C, Coletta RD, Mäkitie AA, Salo T, Almangush A, Leivo I. Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer. Int J Med Inform 2020;136:104068.

[58] Farhadian M, Shokouhi P, Torkzaban P. A decision support system based on support vector machine for diagnosis of periodontal disease. BMC Res Notes 2020;13(1):1-6.

[59] Reyes LT, Knorst JK, Ortiz FR, Brondani B, Emmanuelli B, Guedes RS, Mendes FM, Ardenghi TM. Early childhood predictors for dental caries: A machine learning approach. J Dent Res 2023 (doi: 10.1177/00220345231170535).

[60] Sakai T, Li H, Shimada T, Kita S, Lida M, Lee CW, Nakano T, Yamaguchi S, Imazato S. Development of artificial intelligence model for supporting implant drilling protocol decision making. J Prosthodont Res 2023;67(3):360-365.

[61] Dietterich TG. An experimental comparison of three methods for constructing

ensembles of decision trees: Bagging, boosting, and randomization. Mach Learn 2000;40(2):139-57.

[62] Zhang ZY, Shi JC, Yu TY, Santomauro A, Gordon A, Gou JH, Wu DZ. Predicting flexural strength of additively manufactured continuous carbon fiber-reinforced polymer composites using machine learning. J Comput Inf Sci Eng 2020;20(6):061015.

[63] Kern C, Klausch T, Kreuter F. Tree-based machine learning methods for survey research. Surv Res Methods 2019;13(1):73-93.

[64] Song YQ, Lindsay J, Zhao Y, Nasiri A, Louis SY, Ling J, Hu M, Hu JJ. Machine Learning based prediction of noncentrosymmetric crystal materials. Comp Mater Sci 2020; 183:109792.

[65] Domingos P. A few useful things to know about machine learning. Commun ACM 2012;55(10):78-87.

[66] Pomberger A, McCarthy AAP, Khan A, Sung S, Taylor CJ, Gaunt MJ, Colwell L, Walz D, Lapkin AA. The effect of chemical representation on active machine learning towards closed-loop optimization. React Chem Eng 2022;7(6):1368-79.

[67] Esaki T, Yonezawa T, Yamazaki D, Ikeda K. Prediction models for fraction of absorption and membrane permeability using mordred descriptors. Chem-Bio Inform J 2022;22:46-54.

[68] Moriwaki H, Tian YS, Kawashita N, Takagi T. Mordred: a molecular descriptor calculator. J Cheminformatics 2018;10(1):1-14.

[69] Yamashita T, Sato N, Kino H, Miyake T, Tsuda K, Oguchi T. Crystal structure prediction accelerated by Bayesian optimization. Phys Rev Mater

2018;2(1):013803.

[70] Seko A, Hayashi H, Nakayama K, Takahashi A, Tanaka I. Representation of compounds for machine-learning prediction of physical properties. Phys Rev B 2017;95(14):144110.

[71] Yao J, Li J, Wang Y, Huang H. Comparison of the flexural strength and marginal accuracy of traditional and CAD/CAM interim materials before and after thermal cycling. J Prosthet Dent 2014;112(3):649-57.

[72] Nguyen JF, Migonney V, Ruse ND, Sadoun M. Properties of experimental urethane dimethacrylate-based dental resin composite blocks obtained via thermo-polymerization under high pressure. Dent Mater 2013;29(5):535-41.

[73] Asmussen E, Peutzfeldt A. Influence of UEDMA BisGMA and TEGDMA on selected mechanical properties of experimental resin composites. Dent Mater 1998;14(1):51-6.

[74] Sideridou I, Tserki V, Papanastasiou G. Effect of chemical structure on degree of conversion in light-cured dimethacrylate-based dental resins. Biomaterials 2002;23(8):1819-29.

[75] Cornelio RB, Wikant A, Mjosund H, Kopperud HM, Haasum J, Gedde UW, Örtengren UT. The influence of bis-EMA vs bis GMA on the degree of conversion and water susceptibility of experimental composite materials. Acta Odontol Scand 2014;72(6):440-7.

[76] Sideridou I, Tserki V, Papanastasiou G. Effect of chemical structure on degree of conversion in light-cured dimethacrylate-based dental resins. Biomaterials 2002;23(8):1819-29.

[77] Dickens SH, Stansbury JW, Choi KM, Floyd CJE. Photopolymerization kinetics of methacrylate dental resins. Macromolecules 2003;36(16):6043-53.

[78] Szczesio-Wlodarczyk A, Domarecka M, Kopacz K, Sokolowski J, Bociong K. An evaluation of the properties of urethane dimethacrylate-based dental resins. Materials 2021;14(11):2727.

[79] Schabbach LM, Bondioli F, Fredel MC. Color prediction with simplified Kubelka-Munk model in glazes containing Fe2O3-ZrSiO4 coral pink pigments. Dyes Pig 2013;99(3):1029-35.

[80] Krivak R, Hoksza D, Skoda P. Improving quality of ligand-binding site prediction with Bayesian optimization. Ieee Int C Bioinform 2017:2278-9.

[81] Griffiths RR, Hernandez-Lobato JM. Constrained Bayesian optimization for automatic chemical design using variational autoencoders. Chem Sci 2020;11(2):577-86.

[82] Schweidtmann AM, Clayton AD, Holmes N, Bradford E, Bourne RA, Lapkin AA. Machine learning meets continuous flow chemistry: Automated optimization towards the Pareto front of multiple objectives. Chem Eng J 2018;352:277-82.

[83] Reker D, Hoyt EA, Bernardes GJL, Rodrigues T. Adaptive optimization of chemical reactions with minimal experimental information. Cell Rep Phys Sci 2020;1(11):100247.

[84] Ranković B, Griffiths R-R, Moss HB, Schwaller P. Bayesian optimisation for additive screening and yield improvements in chemical reactions–beyond one-hot encoding. ChemRxiv 2023 (doi: 10.26434/chemrxiv-2022-nll2j-v3).

[85] Deshwal A, Simon CM, Doppa JR. Bayesian optimization of nanoporous materials.

Mol Syst Des Eng 2021;6(12):1066-86.

[86] Xie YC, Zhang C, Deng H, Zheng BJD, Su JW, Shutt K, Lin J. Accelerate synthesis of metal-organic frameworks by a robotic platform and Bayesian Optimization. Acs Appl Mater Inter 2021;13(45):53485-91.

[87] Agarwal G, Doan HA, Robertson LA, Zhang L, Assary RS. Discovery of energy storage molecular materials using quantum chemistry-guided multiobjective Bayesian Optimization. Chem Mater 2021;33(20):8133-44.

Table 1. Detailed information of the computer-aided design/computer-aided manufacturing resin composite blocks used.

Trade name	Manufacturer	Composition*			Flexural strength (MPa)#
		Monomer	Filler	Filler content (wt%)	
Cerasmart	GC, Tokyo,Japan	UDMA, Bis-MEPP	SiO_2, Barium glass	71	174.2 [34]^
Katana Avencia Block	Kuraray Noritake Dental, Niigata, Japan	UDMA, Methacrylatemonomer	SiO_2, Al_2O_3	62	198.2 [34]^
Katana Avencia P Block	Kuraray Noritake Dental	UDMA	SiO_2, Barium glass,Methacrylate mixed filler	82	270
Shofu Block HC	Shofu, Kyoto, Japan	UDMA, TEGDMA	SiO_2, $ZrSiO_4$, Micro-fumed silica	68	132 [34]^
Shofu Block HC Hard	Shofu	UDMA	SiO_2, $ZrSiO_4$, Micro-fumed silica	76	227 [34]^
KZR-CAD HR2	Yamakin, Osaka, Japan	UDMA, TEGDMA	SiO_2, Al_2O_3, SiO_2-ZrO_2 filler	74	162
Estelite Block	Tokuyama Dental, Ibaraki, Japan	UDMA, TEGDMA	SiO_2, SiO_2-ZrO_2 filler	70	195
Estelite P Block	Tokuyama Dental	UDMA, NPGDMA, Bis-MEPP	SiO_2, SiO_2-ZrO_2 filler	81	257
Brillant Crios	Coltene, Altstätten, Switzerland	TEGDMA, Bis-GMA, Bis-EMA [30]^	SiO_2, Barium glass [30]^	71 [29]^	213 [34]^
Lava Ultimate	3M, St. Paul, MN, USA	UDMA, TEGDMA, Bis-GMA, Bis-EMA [29,31]^	SiO_2, ZrO_2, SiO_2-ZrO_2 filler [29,31]^	79 [5]^	197.7 [34]^
Paradigm MZ 100	3M	TEGDMA,Bis-GMA [32]^	SiO_2, SiO_2-ZrO_2 filler [32]^	85 [32]^	189.7 [36]^
Tetric CAD	Ivoclar Vivadent, Schaan,Liechtenstein	UDMA, TEGDMA, Bis-GMA, Bis-EMA [33]^	SiO2, Barium glass [33]^	71 [33]^	185 [34]^

UDMA: urethane dimethacrylate, Bis-MEPP: 2,2-Bis(4-methacryloxypolyethoxyphenyl)propane,

TEGDMA: triethylene glycol dimethacrylate, NPGDMA: neopentyl glycol dimethacrylate,

Bis-GMA: bisphenol A glycidyl methacrylate, Bis-EMA: ethoxylated bisphenol A-glycol dimethacrylate.

*# The compositional information and flexural strengths were collected from references and manufacturers.

^ The references are shown in the bracket for the corresponding materials.

Table 2. The compositional descriptors used.

No.	Descriptors
1	SiO_2 filler
2	ZrO_2 filler
3	$ZrSiO_4$ filler
4	Micro-fumed silica filler
5	Barium glass filler
6	Al_2O_3 filler
7	Methacrylate mixed filler
8	SiO_2-ZrO_2 filler
9	UDMA
10	Bis-MEPP
11	TEGDMA
12	NPGDMA
13	Bis-GMA
14	Bis-EMA
15	Filler content

Table 3. Different fluctuation ranges for 0 and 1.

Fluctuation range	0	1
0.1	[0,0.1]	[0.95,1.05]
0.2	[0,0.2]	[0.90,1.10]
0.3	[0,0.3]	[0.85,1.15]
0.4	[0,0.4]	[0.80,1.20]
0.5	[0,0.5]	[0.75,1.25]
0.6	[0,0.6]	[0.70,1.30]
0.7	[0,0.7]	[0.65,1.35]
0.8	[0,0.8]	[0.60,1.40]
0.9	[0,0.9]	[0.55,1.45]
1.0	[0,1.0]	[0.50,1.50]

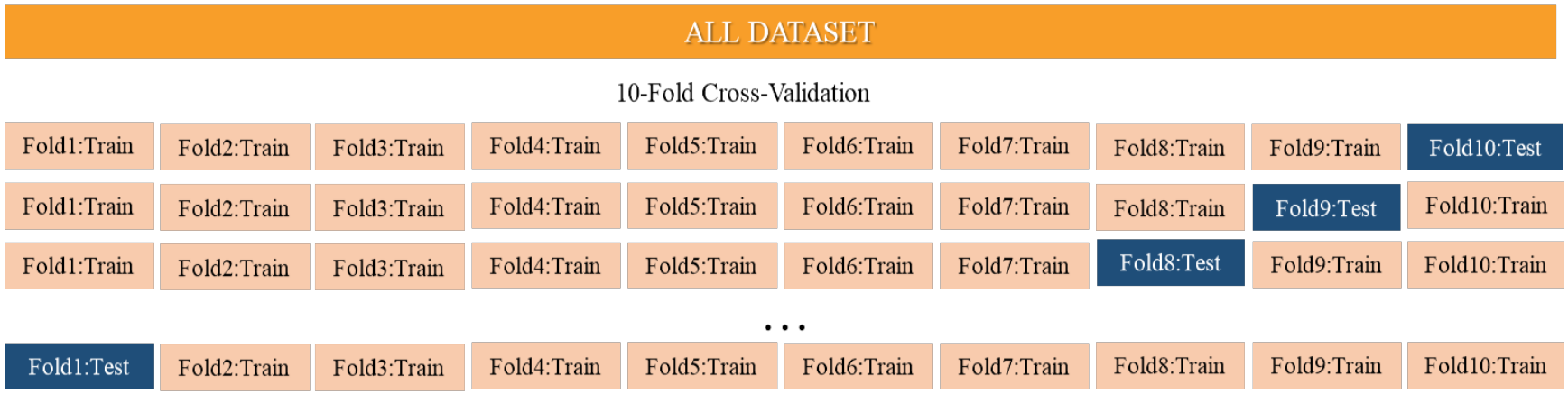


Figure 1. Schematic structure of 10-fold cross-validation

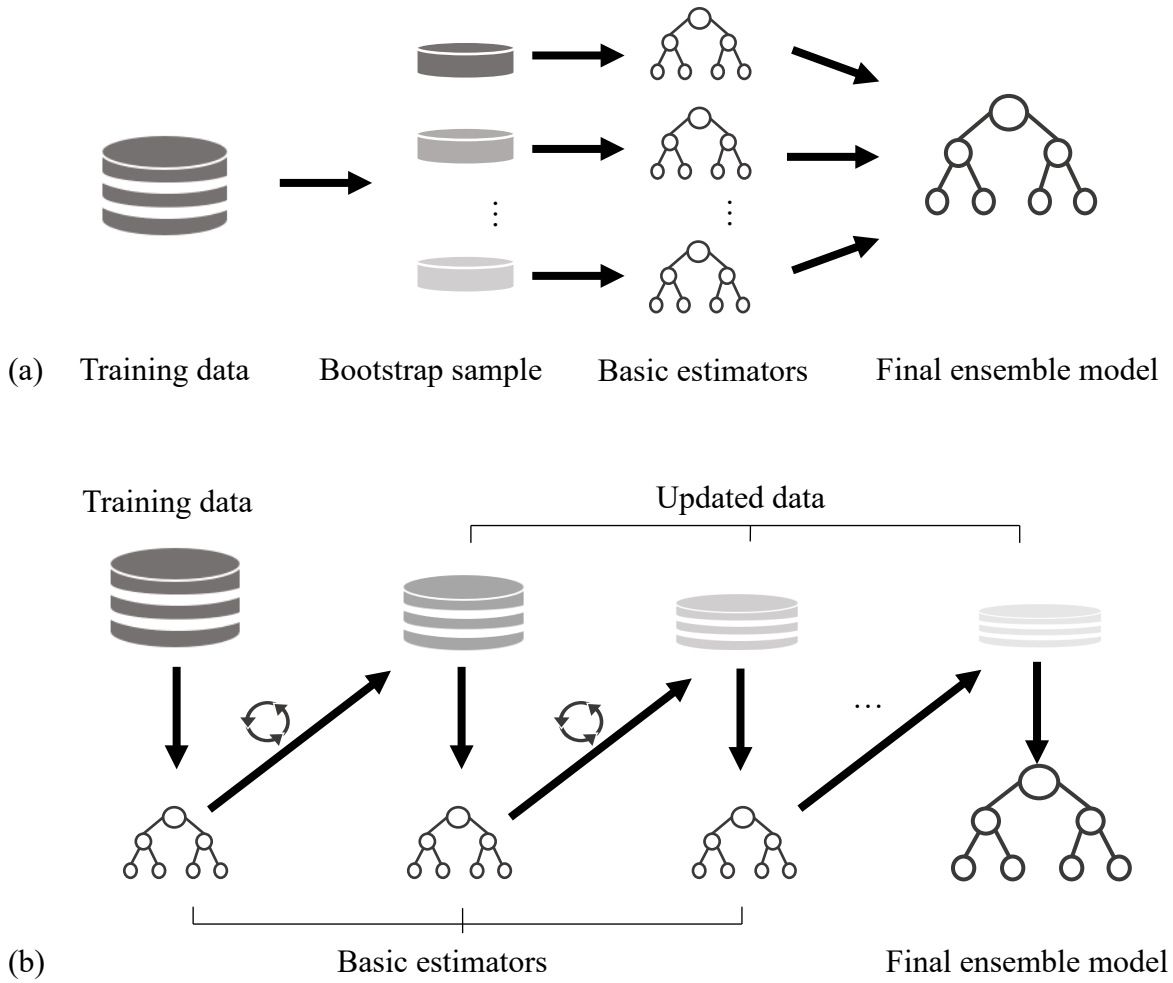


Figure 2. The schematic structure of bagging and boosting.

(a) Bagging: Each basic estimator is independent. (b) Boosting: Each basic estimator is created sequentially based on the former results, trying to correct its shortcomings. The three round arrows represent data updating process.

Table 4. The R^2, *RMSE*, and *MAE* values of different fluctuation range dataset for RF model.

Fluctuation range	RF		
	R^2	*RMSE*	*MAE*
0.1	0.972±0.015	5.199±1.277	4.048±1.043
0.2	0.969±0.009	6.384±1.911	4.283±0.808
0.3	0.967±0.012	6.192±1.900	4.855±1.505
0.4	0.966±0.010	5.767±1.304	4.359±1.030
0.5	0.961±0.037	6.300±1.610	5.034±1.437
0.6	0.957±0.020	6.688±1.891	5.276±1.796
0.7	0.928±0.032	9.173±3.380	5.629±1.364
0.8	0.855±0.058	12.986±4.169	9.227±3.250
0.9	0.708±0.265	16.263±4.623	11.444±2.303
1.0	0.729±0.092	17.450±4.697	13.363±3.029

Table 5. The R^2, *RMSE*, and *MAE* values of different fluctuation range dataset for ET model.

Fluctuation range	ET		
	R^2	*RMSE*	*MAE*
0.1	0.995±0.004	2.172±0.900	1.533±0.710
0.2	0.994±0.004	2.357±0.424	1.810±0.413
0.3	0.989±0.009	3.100±0.545	2.355±0.533
0.4	0.984±0.008	4.077±0.910	3.145±0.623
0.5	0.982±0.009	4.289±0.848	3.344±0.529
0.6	0.968±0.017	5.729±1.711	4.698±1.408
0.7	0.894±0.038	11.178±3.222	7.627±1.604
0.8	0.863±0.037	12.692±3.698	10.301±3.025
0.9	0.873±0.041	11.913±2.158	9.643±1.725
1.0	0.832±0.053	13.647±2.206	10.921±1.316

Table 6. The R^2, *RMSE*, and *MAE* values of different fluctuation range dataset for GBDT model.

Fluctuation range	GBDT		
	R^2	*RMSE*	*MAE*
0.1	0.999±0.001	0.678±0.374	0.435±0.190
0.2	0.991±0.007	2.939±1.701	1.893±0.941
0.3	0.998±0.002	1.081±1.062	0.530±0.389
0.4	0.998±0.001	1.664±0.474	0.781±0.221
0.5	0.984±0.013	3.108±1.078	2.247±0.649
0.6	0.997±0.004	1.183±0.826	0.559±0.218
0.7	0.927±0.058	8.439±5.652	3.918±1.703
0.8	0.828±0.104	13.625±5.798	7.520±3.075
0.9	0.705±0.325	15.423±4.698	10.800±2.163
1.0	0.660±0.333	16.977±4.425	11.550±3.542

Table 7. The R^2, *RMSE*, and *MAE* values of different fluctuation range dataset for XGBoost model.

Fluctuation range	XGBoost		
	R^2	*RMSE*	*MAE*
0.1	0.956±0.029	6.070±1.919	3.805±1.067
0.2	0.955±0.046	6.282±2.649	4.303±1.973
0.3	0.950±0.011	8.023±1.669	5.964±1.176
0.4	0.948±0.010	7.604±0.928	3.771±0.627
0.5	0.954±0.031	7.295±2.016	4.823±1.401
0.6	0.955±0.028	7.330±3.625	5.373±2.585
0.7	0.898±0.055	11.089±4.850	6.489±1.797
0.8	0.730±0.030	18.070±4.639	13.310±4.300
0.9	0.708±0.335	15.404±5.973	11.523±3.663
1.0	0.635±0.211	18.606±2.326	14.236±2.029

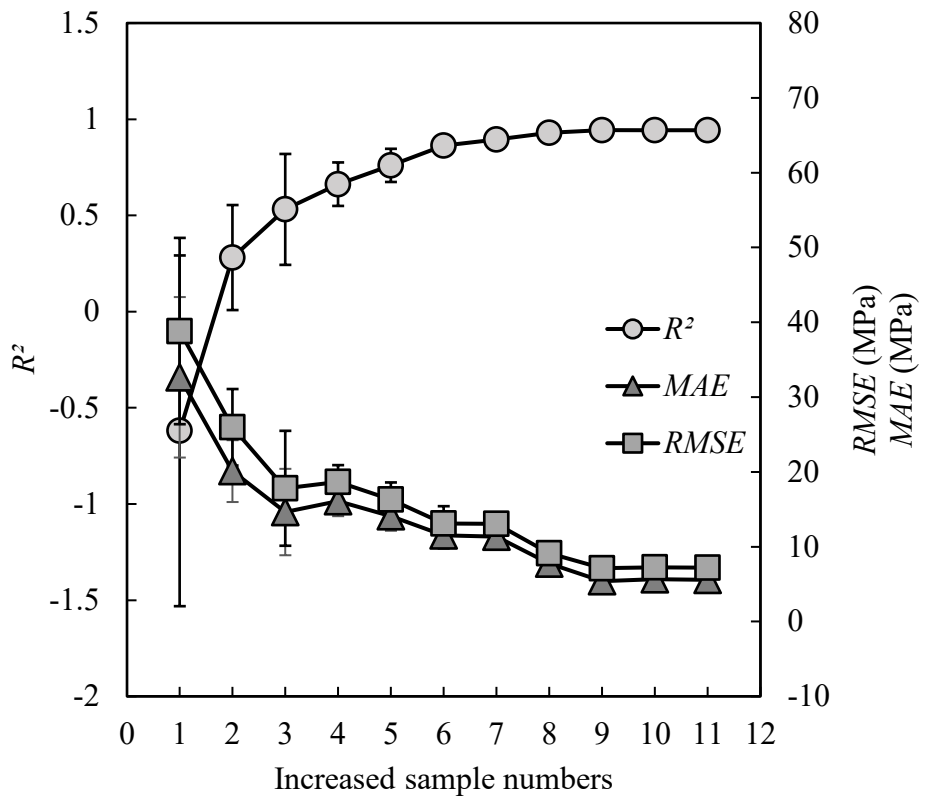


Figure 3. The R^2, *RMSE*, and *MAE* values of different increased sample number dataset for RF model.

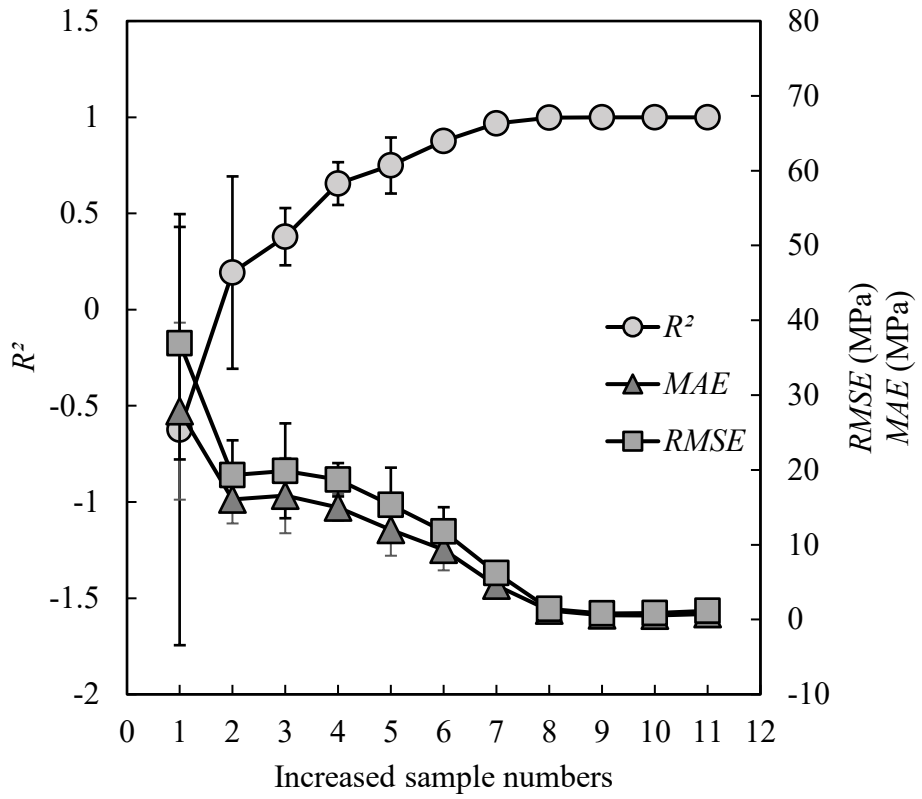


Figure 4. The R^2, *RMSE*, and *MAE* values of different increased sample number dataset for ET model.

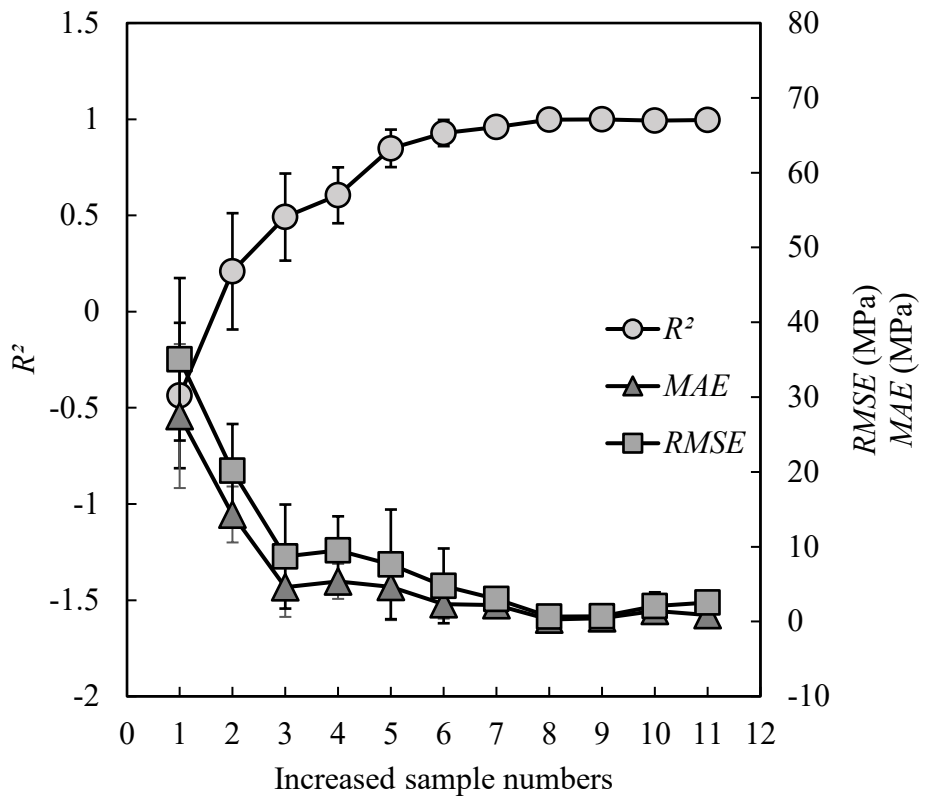


Figure 5. The R^2, *RMSE*, and *MAE* values of different increased sample number dataset for GBDT model.

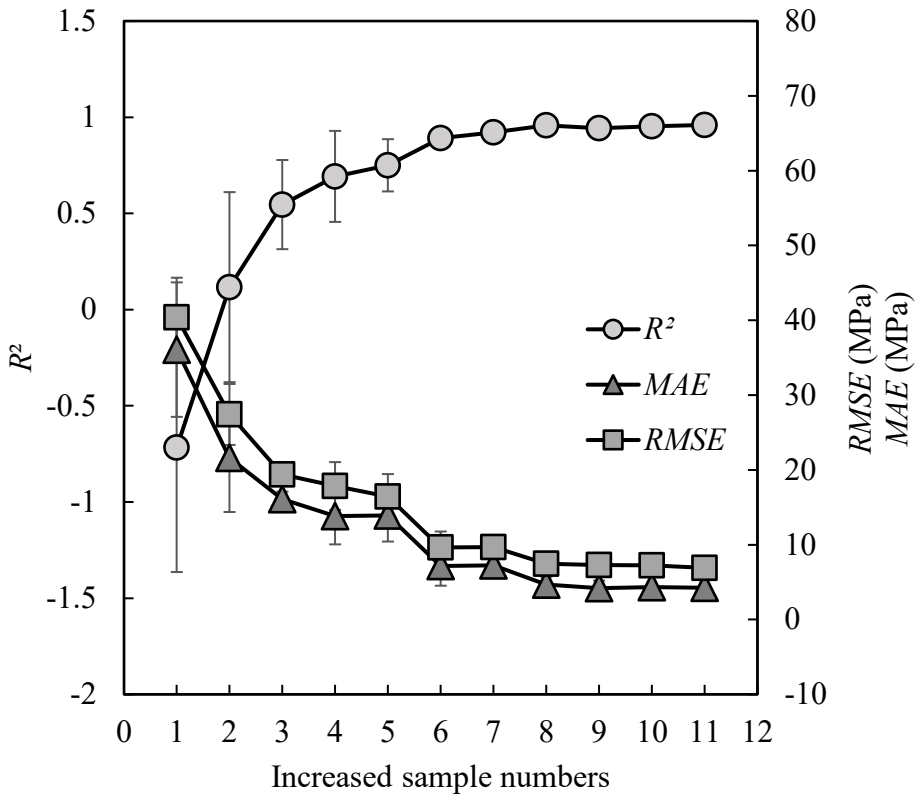


Figure 6. The R^2, *RMSE*, and *MAE* values of different increased sample number dataset for XGBoost model.

Table 8. Tuned hyperparameters for the implemented algorithms.

Models	Tuned hyperparameters
RF	n_estimators=280, max_depth=6, min_samples_leaf=1, min_samples_split=3, max_features=14, random_state=100
ET	n_estimators=41, max_depth=6, min_samples_leaf=1, min_samples_split=2, random_state=3
GBDT	n_estimators=120, max_depth=5, learning_rate=0.04, min_samples_leaf=1, min_samples_split=2, min_weight_fraction_leaf=0.04, random_state=4
XGBoost	n_estimators=127, max_depth=3, learning_rate=0.55, booster='dart', min_child_weight=6, reg_alpha=0.01, colsample_bytree=1, colsample_bylevel=1, random_state=3

Table 9. Prediction performance of the four ML models.

Model	*RMSE* (MPa)	*MAE* (MPa)	R^2	Percent of predictions within relative error of				
				1%	2.5%	5%	10%	15%
RF	5.643	4.753	0.947	20.83	58.33	91.67	100.00	100.00
ET	1.240	0.725	0.997	95.83	95.83	100.00	100.00	100.00
GBDT	1.179	0.761	0.998	91.67	100.00	100.00	100.00	100.00
XGBoost	6.639	4.142	0.927	29.17	75.00	91.67	95.83	100.00

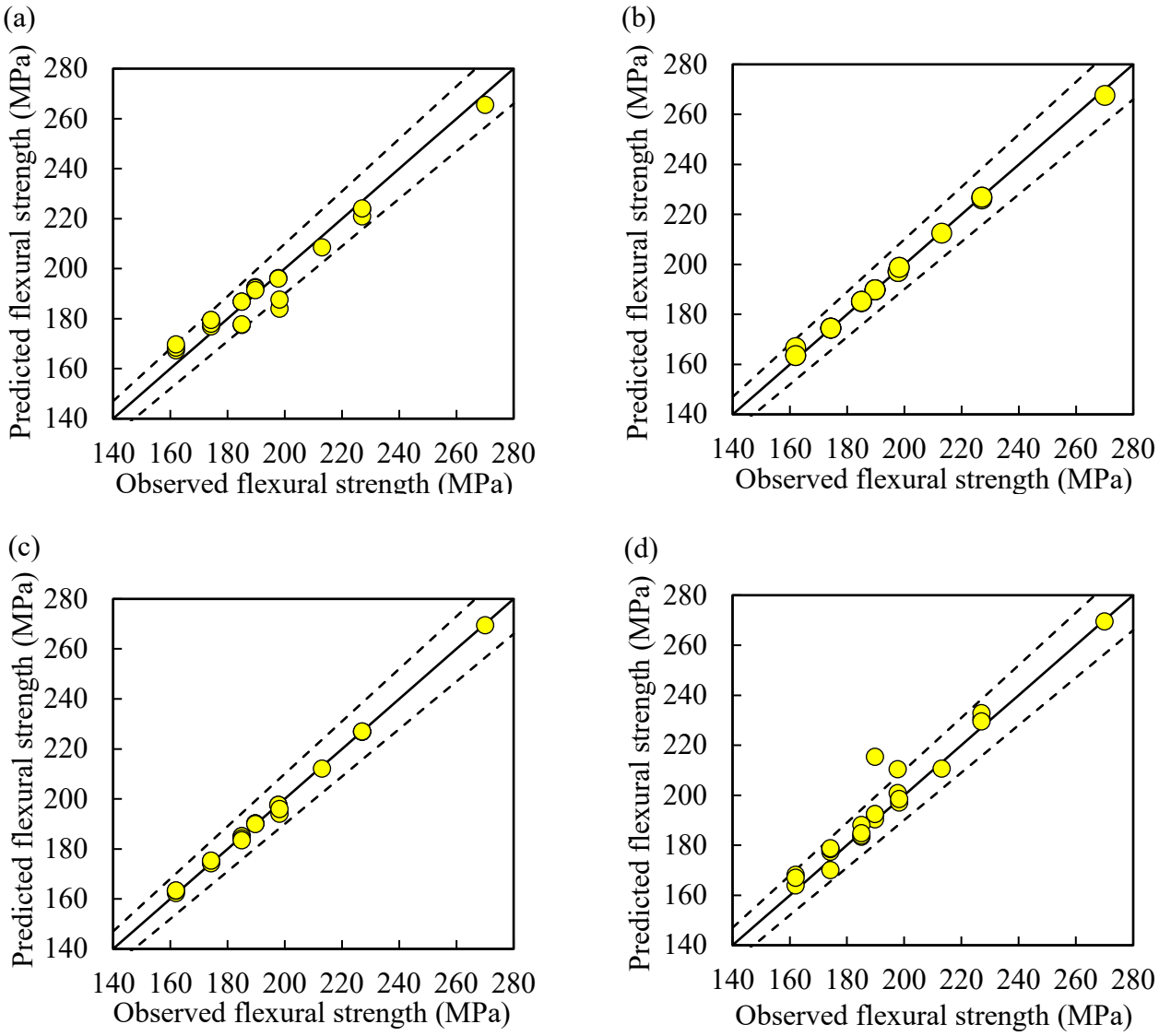


Figure 7. Prediction performance of the ML models on the test set.

The predicted flexural strengths from the four ML models are plotted versus the *in vitro* three-point bending test result (observed value). (a) RF model; (b) ET model; (c) GBDT model; (d) XGBoost model. The data points within the region of two black dashed lines have relative errors less than 5%.

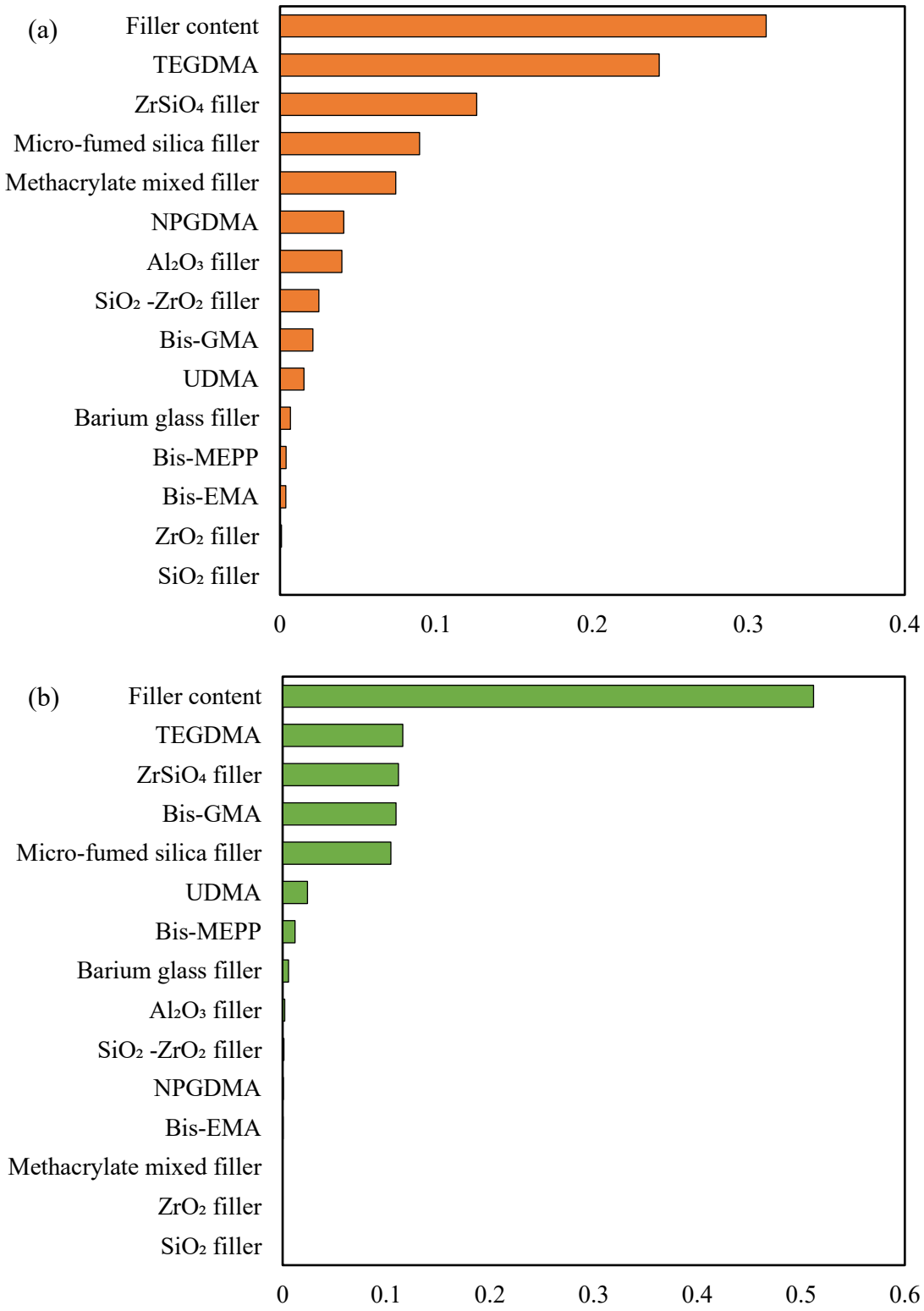


Figure 8. Feature importance for each composition calculated by ET and GBDT models.

(a) ET model; (b) GBDT model.

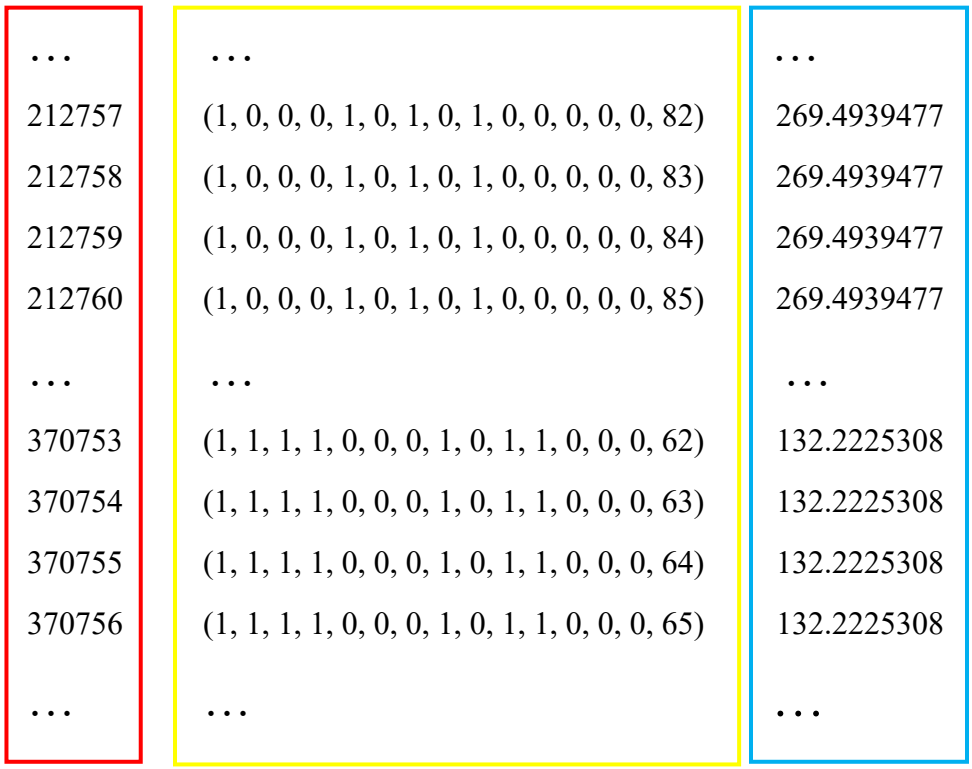


Figure 9. Part of the original output of exhaustive search by GBDT model.

The original output of the exhaustive search contains three parts: the combination numbers, the corresponding compositions and the predicted flexural strengths. The combination numbers are from 1 to 393,216, and some of these numbers are displayed in the red square. For each combination, it represents a specific composition for the CAD/CAM RCB that was formed by 0 ,1 and filler content (62-85) as shown in the yellow square. The predicted flexural strength for each composition was shown in the blue square.

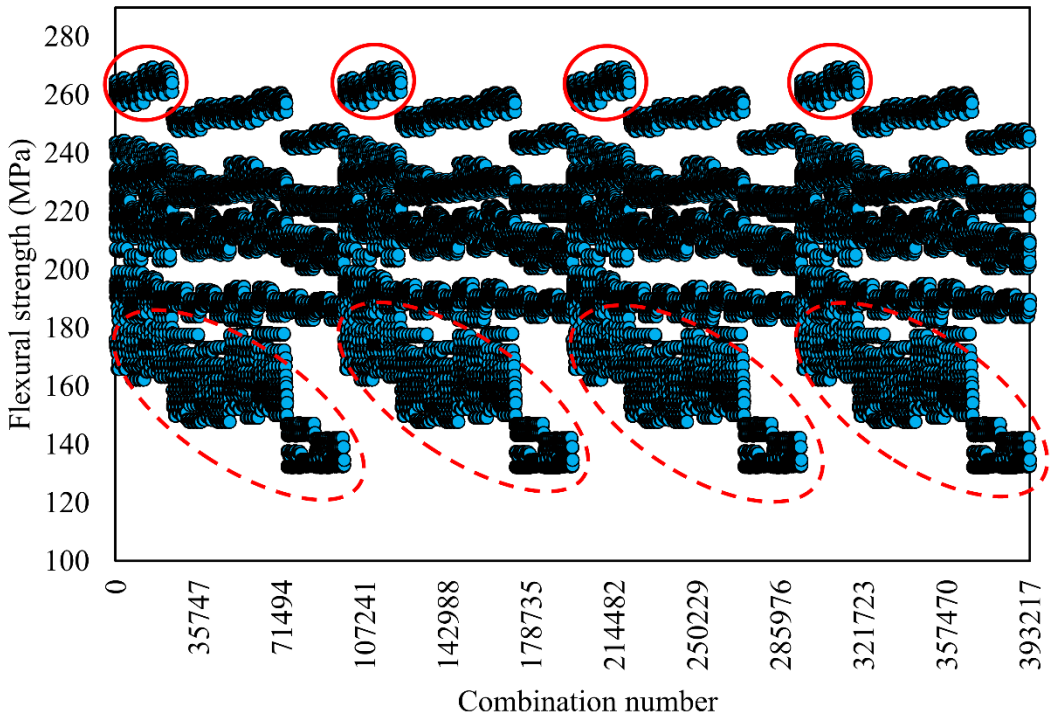
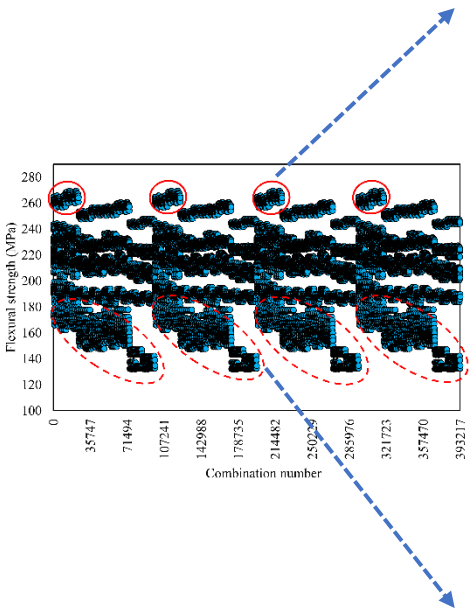


Figure 10. Flexural strengths of 2^{14} × 24 combinations of descriptors predicted by the GBDT model using an exhaustive search.

The bottom of the figure shows four high-density areas of points as shown by the red dotted circle, where the predictions varied from 132.2 MPa to 180.1 MPa, containing around 93,135 possible combinations. By contrast, the high predictions group (shown in red circle) ranged from 256.6 MPa to 269.5 MPa, containing 4,527 possible combinations.



Filler	Frequency(%)
SiO_2	100.0
ZrO_2	51.1
$ZrSiO_4$	27
Micro-fumed silica	1.1
Barium glass	61.6
Al_2O_3	51.2
Methacrylate mixed filler	54.2
SiO_2-ZrO_2 filler	46.3

Monomer	Frequency(%)
UDMA	53.9
Bis-MEPP	49.3
TEGDMA	0
NPGDMA	50.7
Bis-GMA	0
Bis-EMA	53.9
Filler content (wt%)	
81-85	

Filler	Frequency(%)
SiO_2	100.0
ZrO_2	50.1
$ZrSiO_4$	59.2
Micro-fumed silica	60.6
Barium glass	51.3
Al_2O_3	50.1
Methacrylate mixed filler	50.1
SiO_2-ZrO_2 filler	50.2

Monomer	Frequency(%)
UDMA	52.1
Bis-MEPP	53.5
TEGDMA	53.5
NPGDMA	50.7
Bis-GMA	50.6
Bis-EMA	50.7
Filler content (wt%)	
62-69	

Figure 11. The composition differences in the higher prediction group and lower prediction group predicted by GBDT model.

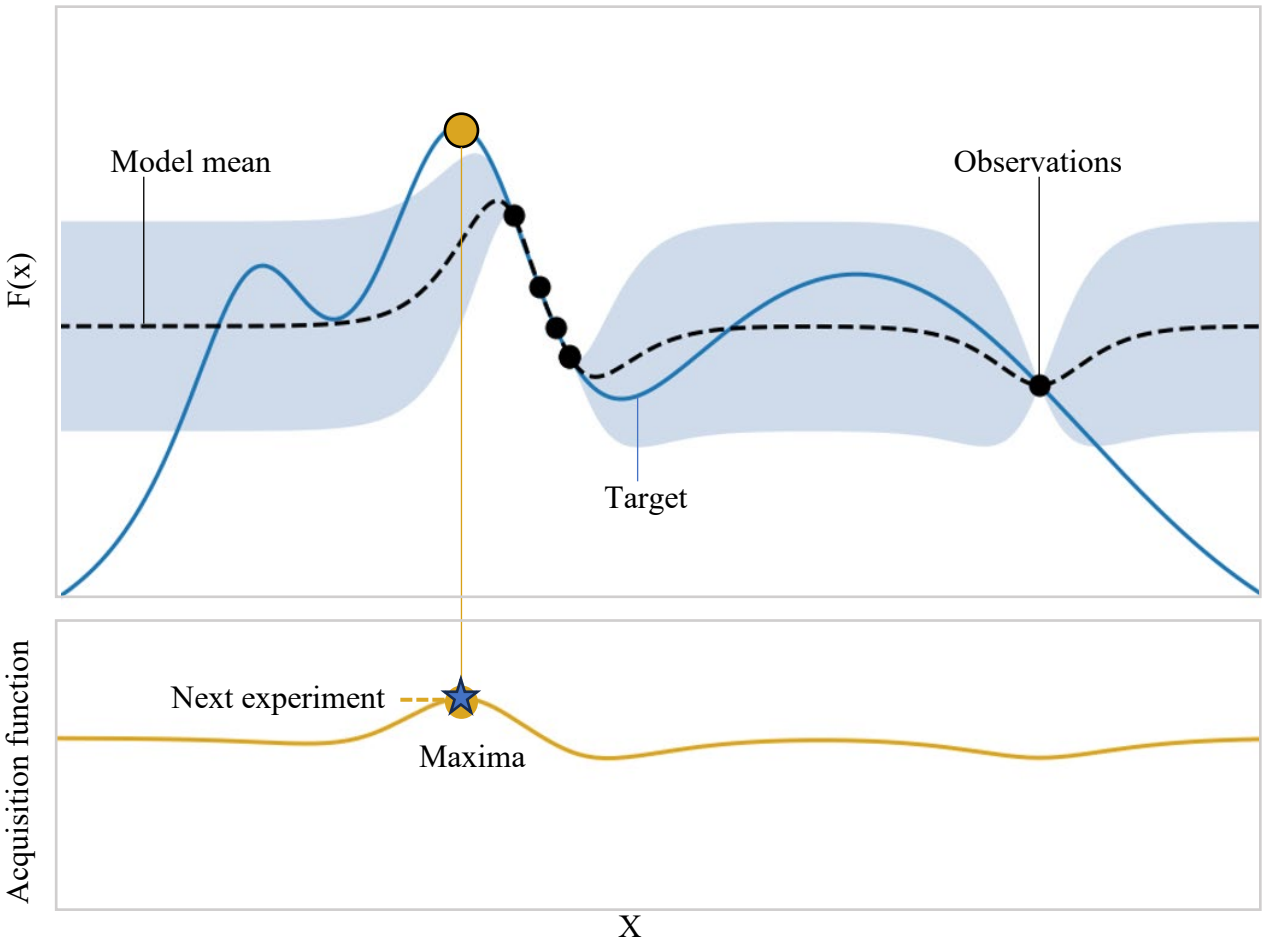


Figure 12. Graphical overview of Bayesian optimization.

One-dimensional example depicting a Gaussian process surrogate model fitted to initialized data and the acquisition function is maximized to select the next configurations. The surrogate model is plotted as the model mean, with the shaded region showing 95% confidence interval. The horizontal axis X could represent the configurations ("chemical composition" in this study). The F(x) could represent the target property that needs to be optimized ("flexural strength" in this study).

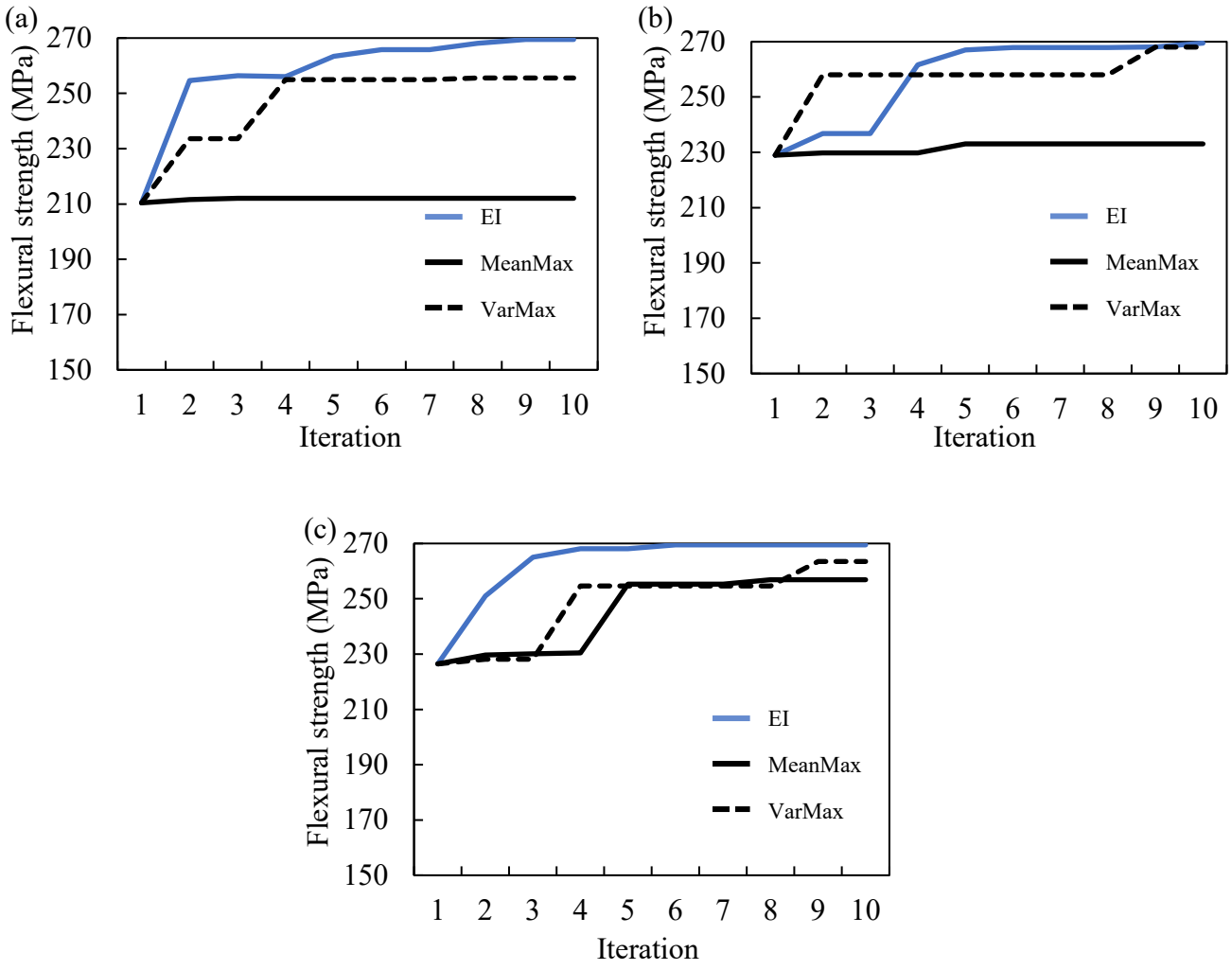


Figure 13. Acquisition function performance of three times initialization on finding the highest flexural strength and the corresponding compositions predicted by GBDT model.

(a) First initialization; (b) Second initialization; (c) Third initialization.

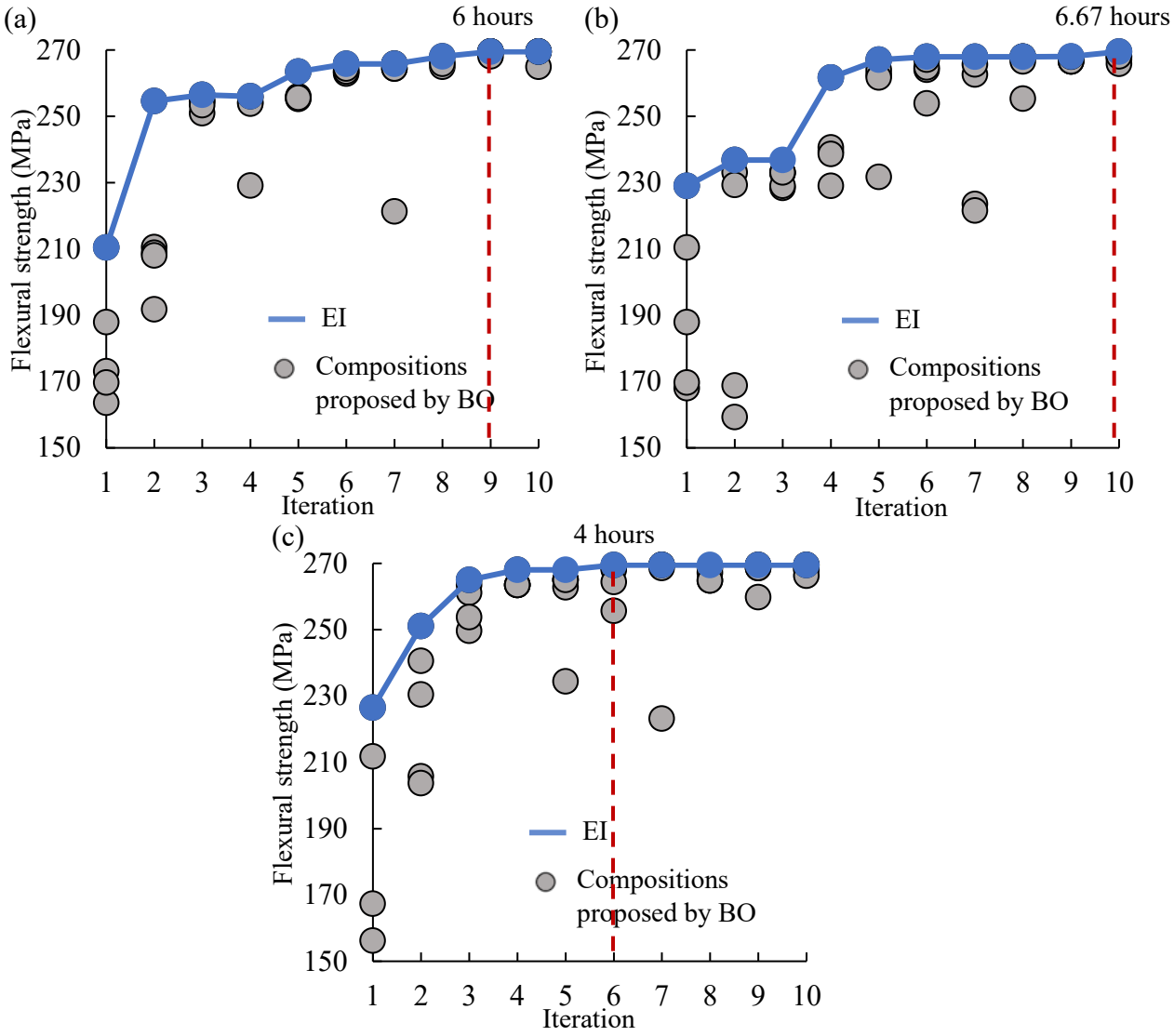


Figure 14. Detailed EI's performance of three times initialization on finding the highest flexural strength and the corresponding compositions predicted by GBDT model.

(a) First initialization: EI proposed one composition led to the highest flexural strength (269.5 MPa) on 9th iteration using 6.00 hours; (b) Second initialization: EI proposed one composition led to the highest flexural strength (269.5 MPa) on 10th iteration using 6.67 hours; (c) Third initialization: EI proposed one composition led to the highest flexural strength (269.5 MPa) on 6th iteration using 4.00 hours .

GBDT model predictions from BO

Maximum predicted flexural strength: 269.5 MPa

Filler: SiO_2, barium glass, methacrylate mixed filler, Al_2O_3

Monomer: UDMA

Filler content: 85wt%

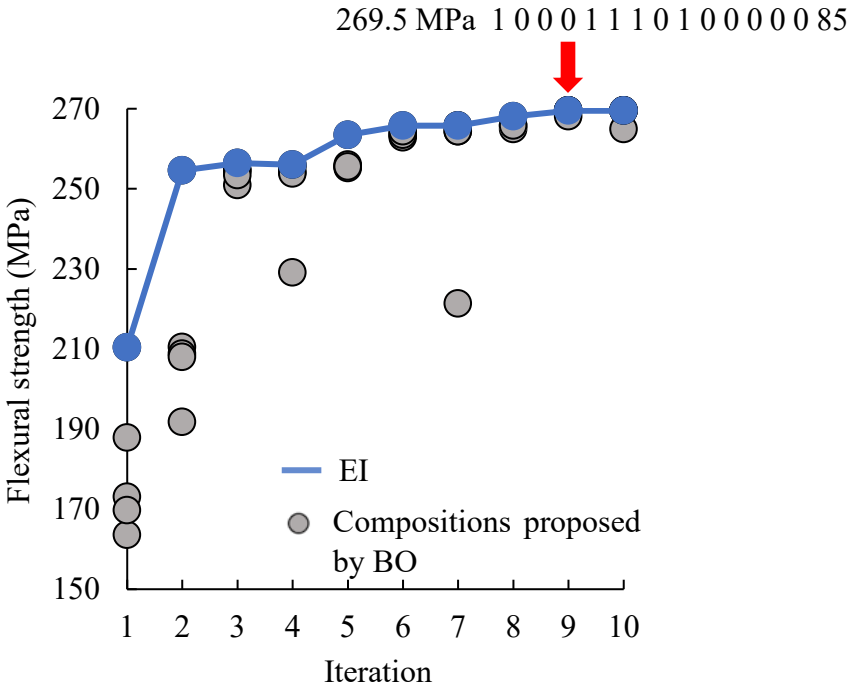


Figure 15. An example from first initialization of EI.

BO proposed a candidate composition in the 9th iteration, and the GBDT model predicted this composition, which resulted in a flexural strength of 269.5 MPa. By checking the descriptors' name, the composition was illustrated in the upper panel.

GBDT model predictions from exhaustive search

Maximum predicted flexural strength: 269.5 MPa

Filler: SiO_2, barium glass, methacrylate mixed filler, with or without Al_2O_3

Monomer: UDMA

Filler content: 82wt% to 85wt%

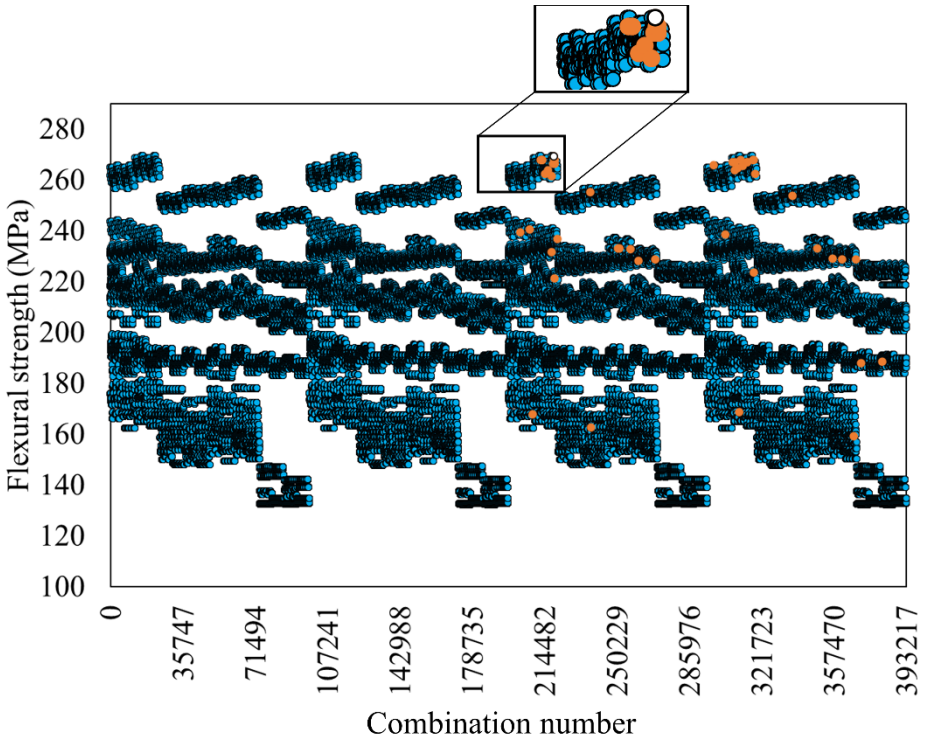


Figure 16. Compositions proposed by BO of 10 iterations in the first initialization.

The proposed compositions were demonstrated in the exhaustive search plot (orange and white dots). The white dot represented the composition showed in Figure 15, which was contained in the concluded compositions from exhaustive search as shown in the upper panel.