

Article

# The Experimental Process Design of Artificial Lightweight Aggregates Using an Orthogonal Array Table and Analysis by Machine Learning

Young Min Wie <sup>1</sup>, Ki Gang Lee <sup>1</sup>, Kang Hyuck Lee <sup>2</sup>, Taehoon Ko <sup>3</sup> and Kang Hoon Lee <sup>4,\*</sup>

<sup>1</sup> Department of Materials Engineering, Kyonggi University, Suwon 16227, Korea; supreme98@kyonggi.ac.kr (Y.M.W.); gglee@kyonggi.ac.kr (K.G.L.)

<sup>2</sup> Center for Built Environment, Sungkyunkwan University, Suwon 16419, Korea; diappong@hanyang.ac.kr

<sup>3</sup> Department of Medical Informatics, The Catholic University of Korea, Seoul 06591, Korea; thko@catholic.ac.kr

<sup>4</sup> Department Civil & Environmental Engineering, Hanyang University, Seoul 04763, Korea

\* Correspondence: diasyoung@hanyang.ac.kr; Tel.: +82-31-249-9774; Fax: +82-31-244-6300

Received: 3 November 2020; Accepted: 4 December 2020; Published: 7 December 2020



**Abstract:** The purpose of this study is to experimentally design the drying, calcination, and sintering processes of artificial lightweight aggregates through the orthogonal array, to expand the data using the results, and to model the manufacturing process of lightweight aggregates through machine-learning techniques. The experimental design of the process consisted of  $L_{18}(3^6 6^1)$ , which means that  $3^6 \times 6^1$  data can be obtained in 18 experiments using an orthogonal array design. After the experiment, the data were expanded to 486 instances and trained by several machine-learning techniques such as linear regression, random forest, and support vector regression (SVR). We evaluated the predictive performance of machine-learning models by comparing predicted and actual values. As a result, the SVR showed the best performance for predicting measured values. This model also worked well for predictions of untested cases.

**Keywords:** orthogonal array experiment design method; lightweight aggregate; support vector regression; machine learning; sintering process

## 1. Introduction

Aggregates are the second most commonly consumed raw material by humanity after water. Therefore, the aggregate industry requires a highly strategic approach in terms of the supply of raw materials for construction and to protect the environment. [1] According to an article in the Korea Construction Newspaper, in 2019 the amount of aggregate used in Korea stood at approximately 200 million tons per year, half of which is mined directly [2]. This presents a major environmental burden and leads to the depletion of natural aggregate resources. Accordingly, in recent years, research focusing on artificial lightweight aggregates has been actively conducted from the viewpoint of environmental protection and waste recycling. Various wastes have been studied, including coal fly ash [3,4], polymer waste [5], sewage sludge [6,7], and waste glass [8,9]. However, these studies focus on the waste itself rather than the manufacturing process of aggregates, and there are relatively few studies on the manufacturing process of lightweight aggregates.

A few manufacturing methods for artificial lightweight aggregate have been considered, with most common being firing in a rotary kiln using natural materials such as expanded clay [10,11]. Numerous studies of firing processes for artificial lightweight aggregates have been conducted. In laboratory-scale research, methods such as rapid sintering [12,13], two-stage sintering [14], and normal sintering are most commonly reported [15,16]. Each experimental result differs slightly

depending on conditions, but it was commonly noted that the calcination section must be shortened to form internal pressure. The importance of internal pressure and viscous behavior for the bloating of artificial lightweight aggregates has recently attracted attention. According to Dondi et al. [17], the foaming properties of an aggregate should be explained not only in terms of the chemical components used but also in terms of the internal pressure and viscous behavior. In a study by Moreno-Maroto et al. [18], viscous behavior was found to be very important for the bloating of aggregates, and only a very small amount of gas is involved in bloating. Moreover, in work by Wie et al. [19], an orthogonal array experimental design was utilized to find the optimal conditions for the bloating activation of aggregates under a normal sintering condition. It was shown that for the bloating of an aggregate, the calcination section should be short and the proper time should be maintained in the calcination section for viscous behavior. This is in line with existing theory [17]. In addition, a prediction of the physical properties using an orthogonal array experiment design was made, and the predicted value was found to be relatively suitable. However, although it was possible to find the optimal conditions and make local predictions under certain experimental conditions using the orthogonal array experiment design, it was difficult to predict conditions that were not tested at all within the experimental range. Additionally, the tendency of the experimental results was not defined uniformly. Therefore, additional techniques are needed to predict untested parts rationally and to define trends clearly according to process changes.

Currently, modeling techniques using machine learning are used in various fields, such as agriculture [20], biology [21], chemistry [22,23], and climate prediction [24]. Regarding construction materials, many studies have been conducted to predict the physical properties of concrete [25–30]. However, these studies require more than 200 sizable instances of data from the literature, making it very difficult to obtain with a typical experiment. For this reason, most studies rely on modeling via the collection of other studies. This is an advantageous way to define an individual variable academically and create an experimental formula. However, this method is not practical to be applied directly to industrial sites because it is difficult to simultaneously apply various variables of industrial sites and individual conditions of all sites cannot be considered. In order to overcome these limitations and create a model that can be practically utilized in the industrial field, it is necessary to model using reliable data under relatively controlled conditions with various variables. The experimental design method using an orthogonal array table is very practical because it can expand a large amount of data with relatively few experiments and, through this, influences on a wide variety of variables can be simultaneously applied. In addition, data expansion by an orthogonal array table can provide a sufficient amount of data required for modeling. In addition, since the machine-learning technique analyzes data with a computer and presents the optimal predicted value, it is expected that the use of the two techniques mentioned above will be very practical in the industrial field.

Therefore, in this study, in order to model the firing process of the artificial lightweight aggregates, data were expanded by the orthogonal array table, and modeled using machine-learning techniques by the expanded data. The applied techniques are linear regression, random forest, and support vector regression techniques, and these are the most used techniques in analysis through machine learning. The suitability of the models was evaluated according to each evaluation index, and the predicted result through the model and the actual test result were compared and analyzed. By verifying the three models and comparing them with actual results, we confirmed the effectiveness of data expansion using an orthogonal array table experimental design, and reviewed the most suitable machine-learning techniques for this.

## 2. Experiment Method

### 2.1. Experimental Variables and Experimental Design

In order to separate the variables according to the unit process of lightweight aggregate and to facilitate data expansion, the experiment was designed as follows. In the experiment, the unit process

was defined for each temperature section. The drying process is defined as the room temperature (r.t.) ~600 °C section. The drying section was divided into r.t. ~300 °C and 300~600 °C again, to distinguish between the evaporation section of water and the section where the combustion of organic matter begins. The calcination process is defined as the 600 °C~bloating activation temperature section, and the bloating activation temperature soaking section is defined as the bloating start and activation section. Table 1 shows the experimental variables. The design utilized an orthogonal array experiment table with the indicated variables. After conducting the designed experiment, the density of the aggregates produced was measured according to KS F 2503 [31].

**Table 1.** Processing variables by item.

Process	Process Temp. (°C)	Variables
Drying and preheating	25 ~ 300 °C	20, 40, 60 min
	300 ~ 600 °C	20, 40, 60 min
Calcination	600 °C ~ B.S.A. temp.	20, 40, 60 min
Bloating Start and Activation (B.S.A.)	1180, 1200, 1220 °C	Time 0, 5, 10, 15, 20, 40 min

## 2.2. Data Expansion by Means of an Orthogonal Array Experiment Design

Full factor design requires a lot of experimentation because in real industry it is usually composed of a large number of factors. The method of selecting a limited number of experiments is called a partial factor experiment, and although this method is widely known, general guidelines for application and analysis have not been established. The orthogonal array table approach drastically reduces the amount of experimentation compared to the full factor experimental design, and the use of data is relatively well known [32–36]. In this study, the number of experiments was significantly reduced by using an orthogonal array table, and data comparable to full factor experiments can be obtained through data expansion. Recently, a research method combining an orthogonal order table and various analysis methods has been widely used. In many cases, it was found that the data obtained from the orthogonal array table and the results obtained from the full factor experiment were not significantly different [37–40].

In order to investigate the possibility that the property prediction by the orthogonal array technique will be used in the study, the predicted value of the density based on the results of the orthogonal array experiment was obtained. The prediction for the experiment was conducted using Minitab (Minitab1.9, Minitab Inc., Pennsylvania, PA, USA). The prediction of the orthogonal array test is a result reflecting the increase or decrease when the variables ( $D_i, D_j, D_k, D_l$ ) desired by the experimenter are compared to the total average value ( $D_{tot.}$ ). This proceeds as follows (Equation (1)):

$$D_{pred.} = D_{tot.} + (D_{tot.} - D_i) + (D_{tot.} - D_j) + (D_{tot.} - D_k) + (D_{tot.} - D_l) \quad (1)$$

( $D_{pred.}$ : Predicted density value,  $D_{tot.}$ : Total average density value,  $D_{i,j,k,l}$ : Average value for each variable)

In this experiment, 18 experiments were expanded to 486 by designing an experiment with an orthogonal array table  $L_{18} (3^6 6^1)$ . This is the same as the number of full factors of the designed experiment, and the planned experiment is expanded to  $3^4 \times 6^1$ . This data expansion is possible because the effects of individual variables can be separated in an experiment designed by an orthogonal array table. Because of this characteristic, experiments using an orthogonal sequence table are widely used in the industry [41–43].

## 2.3. Machine-Learning Analysis Using Extended Data

Figure 1 shows the flow of analysis using extended data using an orthogonal array table. Extended data are trained on three regression models: linear regression, random forest, and supporter vector regression (SVR), while cross-validation is performed using the leave-one-out cross-validation

technique. After that, the model’s fit and predictive performance are determined through four evaluation indicators to find the optimal model. Then, the predicted particle density values obtained through the selected optimal regression model were compared with the experimental values to confirm the effectiveness of the machine-learning technique.

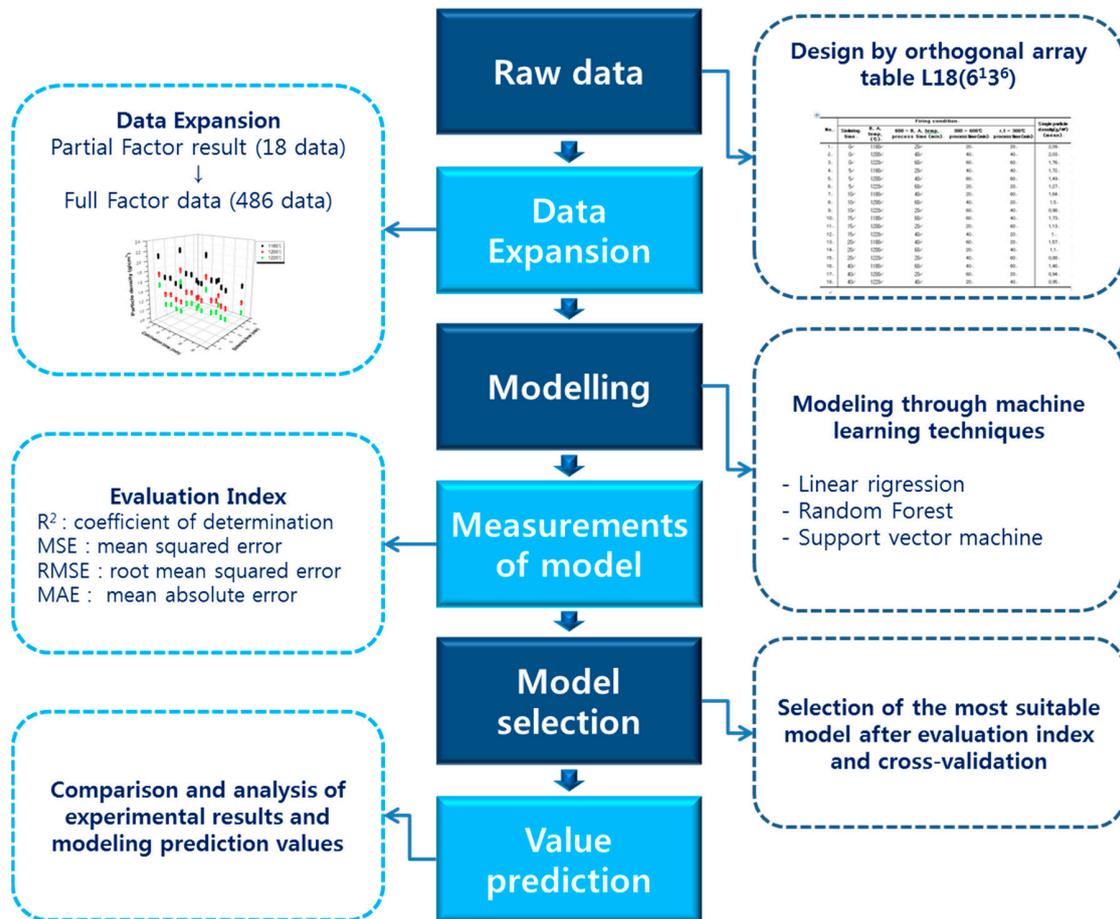


Figure 1. Flow chart for data expansion and machine-learning analysis.

### 3. Machine-Learning Regression Methods

Regression analysis is a set of methods for estimating the relationship between a continuous dependent variable and several independent variables. The functional relationship between the dependent variable  $Y$  and the independent variables  $X_1, X_2, \dots, X_p$  is expressed as the following Equation (2):

$$Y = f(X_1, X_2, \dots, X_p) + \varepsilon \tag{2}$$

where  $\varepsilon$  is called residual, the difference between the observed value of  $Y$  and the predicted value. Regression methods try to find the functional relationship  $f$  in the data.

#### 3.1. Linear Regression

For the given data  $\{y_i, x_{i1}, x_{i2}, \dots, x_{ip}\}_{i=1}^n$ , the linear regression discovers the linear relationship as the following Equation (3):

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i \text{ (for } i = 1, \dots, n) \tag{3}$$

The regression coefficients  $\beta_0, \beta_1, \dots, \beta_p$  can be estimated using the ordinary least-squares (OLS). The least-squares estimate  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$  is defined as the value that minimizes the following Equation (4).

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left\{ y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \right\}^2 \quad (4)$$

Therefore, the predicted value of the dependent variable  $y$  for the new independent variables  $x_1, x_2, \dots, x_p$  can be obtained through the following Equation (5).

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_p x_p \quad (5)$$

### 3.2. Regression Tree

The regression tree is a regression version of decision tree, which is one of the most popular machine learning algorithms. In the process of regression tree training, the entire area of the given data  $\{y_i, x_{i1}, x_{i2}, \dots, x_{ip}\}_{i=1}^n$  is divided into  $M$  areas  $R_1, R_2, \dots, R_M$  and the constant value  $c_m$  is predicted in each area. The tree model that does this is shown below (Equation (6)):

$$f(x) = \sum_{m=1}^M c_m I(x \in R_m). \quad (6)$$

The values of  $c_m$  and  $R_m$  are determined based on the impurity level, and the measurement is the sum of the squared errors  $Q_m(T) = \sum_{i=1}^n (y_i - f(x_i))^2$ . If the given split variable  $x_j$  is continuous, then if the split point is  $s$ ,  $R_1(j, s) = \{x : x_j \leq s\}$  and  $R_2(j, s) = \{x : x_j > s\}$  can be defined. Here, we find the following optimal solution for separation criteria (Equation (7)):

$$\min_{j,s} \left( \min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right) \quad (7)$$

The corresponding solutions having the minimum value for the given  $j$  and  $s$  values are  $\hat{c}_1$  and  $\hat{c}_2$ , which are given as the average  $y_i$  value of the data belonging to  $R_1(j, s)$  and  $R_2(j, s)$ , respectively. After finding the optimal separation criterion  $(j, s)$  through a proper optimization technique, the same process is repeated for both areas. If the new data  $x \in R_m$ , we predict  $\hat{y} = \hat{c}_m$ .

### 3.3. Random Forest

The random forest developed by Breiman [44] takes into account the large variance of decision trees and is a model that uses an ensemble-learning method to learn multiple decision trees randomly. An ensemble is a method that improves prediction accuracy levels by learning multiple models from given data and then synthesizing the prediction results of multiple models.

The random forest algorithm is shown below (Figure 2):

1. Create bootstrap sample  $L^* = \{y_i^*, x_i^*\}_{i=1}^n$  using  $n$  data for training data  $L = \{y_i, x_i\}_{i=1}^n, x_i \in R^p$ .
2. At  $L^*$ , only  $k (\ll p)$  values of independent variables are randomly selected to generate a decision tree.
3. Compute the final predicted value of the random forest by combining the predicted values of the trees. In general, the average value is used.

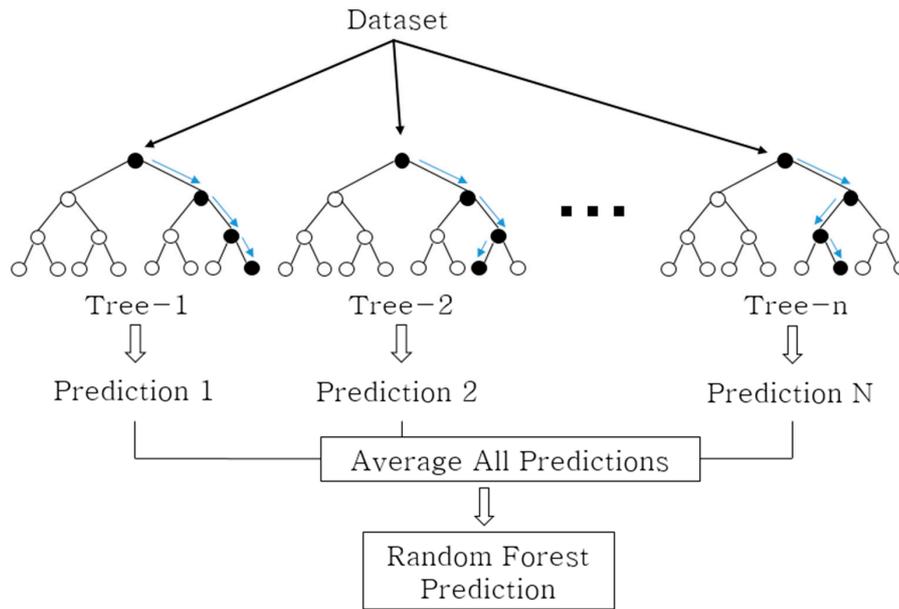


Figure 2. Schematic diagram of the random forest algorithm.

### 3.4. SVR (Support Vector Regression)

Support vector regression (SVR) is a regression version of support vector machine. The purpose of SVR is to find the optimal hyperplane for fitting data instances well. Given the data  $L = \{y_i, x_i\}_{i=1}^n, x_i \in R^m$ , the function of the SVR model proceeds according to Equation (8).

$$f(x) = w^T \cdot x + b, w \in R^m, b \in R \tag{8}$$

This process learns by adding extra variables  $\xi, \xi^*$  that allow errors. The objective function in Equation (9) is optimized so that the norm function of  $w$  is minimized while the deviation between the actual data and the predicted data is within  $\varepsilon$ .

$$\min_{w,b,\xi_i,\xi_i^*} \left( \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi + \xi^*) \right). \tag{9}$$

At this time, the constraint condition is expressed by Equation (10).

$$\begin{aligned} y_i - wx_i - b &\leq \varepsilon + \xi_i \\ b + wx_i - y_i &\leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0, \quad i = 1, 2, \dots, n \end{aligned} \tag{10}$$

The Equation for optimizing the objective function can be solved simply via a dual problem using the Lagrange multiplier and can be expressed as Equation (11) with the Lagrange multiplier.

$$\begin{aligned} L = C \sum_{i=1}^n (\xi_i + \xi_i^*) + \frac{1}{2} \|w\|^2 \\ - \sum_{i=1}^n (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i - y_i + wx_i + b) - \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* + y_i - wx_i - b). \end{aligned} \tag{11}$$

$L$  denotes Lagrangian function and  $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$  are called Lagrangian multipliers; for  $L$ , the differential value of each parameter can be expressed as Equation (12):

$$\frac{\partial L}{\partial w} = 0 \rightarrow w - \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i = 0$$

$$\frac{\partial L}{\partial \xi_i} = 0 \rightarrow c - \alpha_i - \eta_i = 0 \tag{12}$$

Therefore, the objective function of SVM is as expressed as Equation (13).

$$\min_{\alpha_i, \alpha_i^*} \left( \sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) x_i^T x_j \right). \tag{13}$$

The Lagrangian function  $L$  can be expressed as  $\alpha_i, \alpha_i^*$ , and the final prediction model can be obtained using the differential value for  $w$ . (Equation (14))

$$f(x) = w^T \cdot x + b = \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i^T x + b.$$

$$b = y_i - \varepsilon - \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i^T x.$$

$$f(x) = w^T \cdot x + b = \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i^T x + b \tag{14}$$

SVR described above is a linear prediction technique. In order to find a non-linear function as shown in Figure 3, it is necessary to map the given data into a high-dimensional feature space, called kernel space. Several kernel functions are available [45], but the radial basis function (RBF) kernel function is used in this study (Equation (15)):

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2). \tag{15}$$

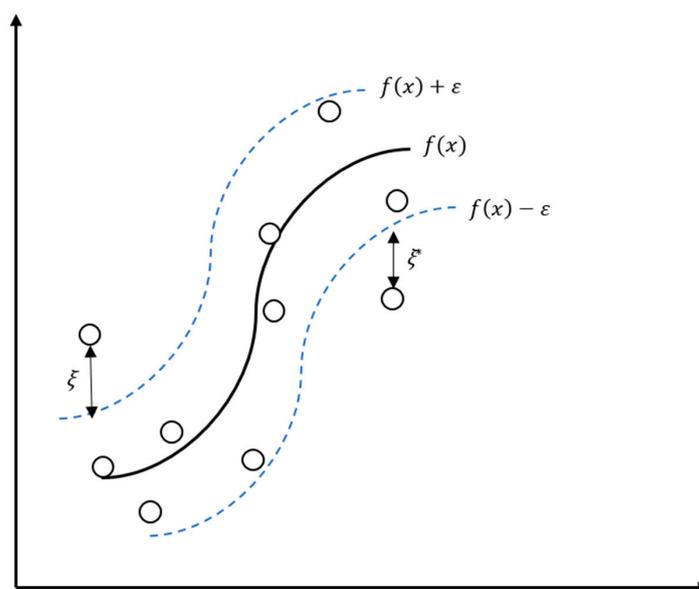


Figure 3. Schematic diagram of non-linear support vector regression.

Therefore, the SVR prediction function, as in Equation (16), can be obtained:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(x_i, x_j) + b \tag{16}$$

## 4. Evaluation of the Performance

### 4.1. Leave-One-Out Cross Validation, (LOOCV)

In general, when evaluating the performance of a predictive model, the model is applied to an unseen test set. In contrast, when using cross-validation, data are divided into several non-overlapping groups, and the model is created by changing the dataset used for training and verification. Cross-validation is widely used as a model evaluation method to compensate for the poor predictability when introducing new data, as the model is overfit to the data used for fitting in machine learning. Leave-one-out cross-validation is a viable means of comparing a value predicted by a model and an actual value with respect to the remaining data after setting the model using the remaining data, apart from one data instance out of the entire dataset. This method is suitable when the number of instances is small.

### 4.2. Performance Evaluation Measures

For an evaluation of the three models used in this paper, the coefficient of determination ( $R^2$ ), the mean squared error (MSE), the root mean squared error (RMSE), and the mean absolute error (MAE) are used.

The  $R^2$  is a measure of the degree to which the estimated linear model fits the given data and shows the explanatory power of the estimated linear model. The  $R^2$  has a value of 1 or less, and it can be said that the closer to 1 the value is, the better the explanatory power of the model becomes.  $y_i$  is the actual target value of the  $i$ -th observation,  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  is the average of  $n$  target values, and  $\hat{y}_i$  is the target value of the  $i$ -th observation predicted by the regression model. The  $R^2$  is expressed by the following Equation (17).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad i = 1, \dots, n \quad (17)$$

The MSE can rank the performances of several models. The average error value is within the range of  $[0, \infty]$ , and the smaller this value is, the better the model performance becomes. Equation (18) for the MSE is as follows:

$$\text{mean squared error} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}. \quad (18)$$

The MSE can effectively rank models, but the difference between the actual and predicted values of the dependent variable is unknown. The RMSE can be used to make a more meaningful interpretation of the model's prediction error. (Equation (19))

$$\text{root mean squared error} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}. \quad (19)$$

Given that the RMSE contains a square term, it tends to overestimate the error because large errors become emphasized. The MAE can be used as an alternative to this situation (Equation (20)).

$$\text{mean absolute error} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}. \quad (20)$$

## 5. Results and Discussion

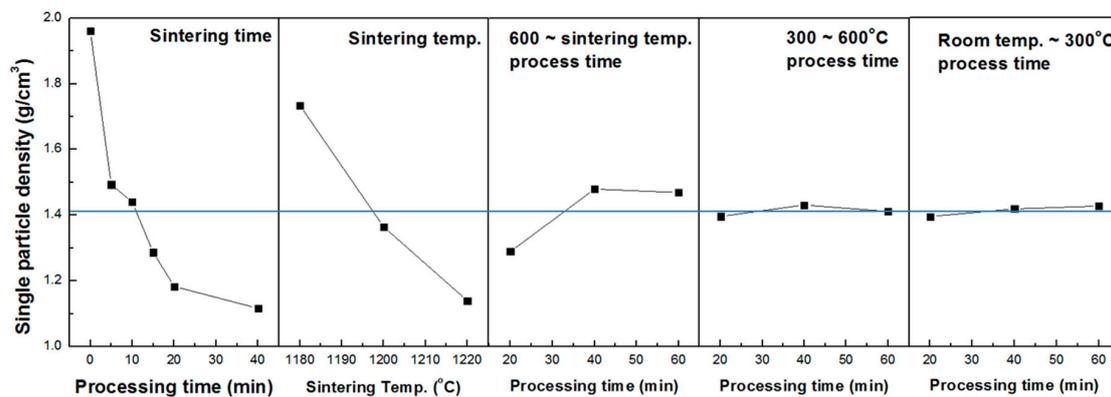
### 5.1. Property Prediction and Data Expansion by the Orthogonal Array Experiment Design

Table 2 shows the orthogonal array experiment design table and single particle density measurement results. The experiment was designed and tested with an orthogonal array table  $L_{18}(3^6 6^1)$ . The three important variables in the experiment were sintering temperature, sintering time, and calcination time. Drying conditions did not significantly affect the properties of the final aggregates.

Figure 4 shows the effect of each variable. It can be seen that the higher the sintering temperature, the lower the density of the aggregate. Moreover, the longer the holding time, the lower the density gradually, but after 20 min, the density no longer decreased. In addition, when the calcination time was 20 min, a lower density level could be obtained.

**Table 2.** Orthogonal array experimental design and the measured density.

No.	Firing Condition					Single Particle Density (g/cm <sup>3</sup> ) (mean)
	Sintering Time	B. S. A. Temp. (°C)	600 ~ B.S.A. Temp. Process Time (min)	300~600 °C Process Time (min)	r.t. ~ 300 °C Process Time (min)	
1	0	1180	20	20	20	2.09
2	0	1200	40	40	40	2.03
3	0	1220	60	60	60	1.76
4	5	1180	20	40	40	1.72
5	5	1200	40	60	60	1.49
6	5	1220	60	20	20	1.27
7	10	1180	40	20	60	1.84
8	10	1200	60	40	20	1.5
9	10	1220	20	60	40	0.98
10	15	1180	60	60	40	1.73
11	15	1200	20	20	60	1.13
12	15	1220	40	40	20	1
13	20	1180	40	60	20	1.57
14	20	1200	60	20	40	1.1
15	20	1220	20	40	60	0.88
16	40	1180	60	40	60	1.46
17	40	1200	20	60	20	0.94
18	40	1220	40	20	40	0.95



**Figure 4.** Analysis result of experiment designed with orthogonal array table.

According to Kōse et al. [46], the generation and development of pores can be divided into three stages.

- (1) Gas is generated inside the aggregates.
- (2) The generated gas creates pressure, the pore walls are destroyed by the pressure, and the pores merge.
- (3) The pores grow due to the pressure difference between small pores and large pores.

At this time, in the initial stage, the processes of (1) and (2) take place, afterwards, it has been found that the process of (3) becomes dominant.

The basic assumption of this theory is that in order to ensure bloating, the aggregate must show viscous behavior. As part of this viscous behavior, the pressure inside changes the pore structure inside. For acidic clay, it was confirmed that the formation of internal pressure is mainly caused by dehydration of crystalline water [19]. When the calcination time is short, because the aggregate reaches the foaming activation temperature before the dehydration of the crystalline water reaction is

completed, the generated gas contributes to foaming by forming pressure inside. It was also found that the higher the bloating activation temperature, the lower the particle density. This occurs because as the sintering temperature increases, the amount of the liquid phase in the aggregate increases; accordingly, the viscous behavior of the aggregate increases. If the time is kept too short at the expansion activation temperature, it is not possible to provide the viscous behavior conditions for the agglomerate expansion, so an adequate holding time is required. It was found that the change in density was saturated at the holding time of 20 min or more.

This result was expanded to 486 data instances, and the modeling of process variables was performed using this data. Figure 5 shows the results of data expansion through the orthogonal array experiment design.

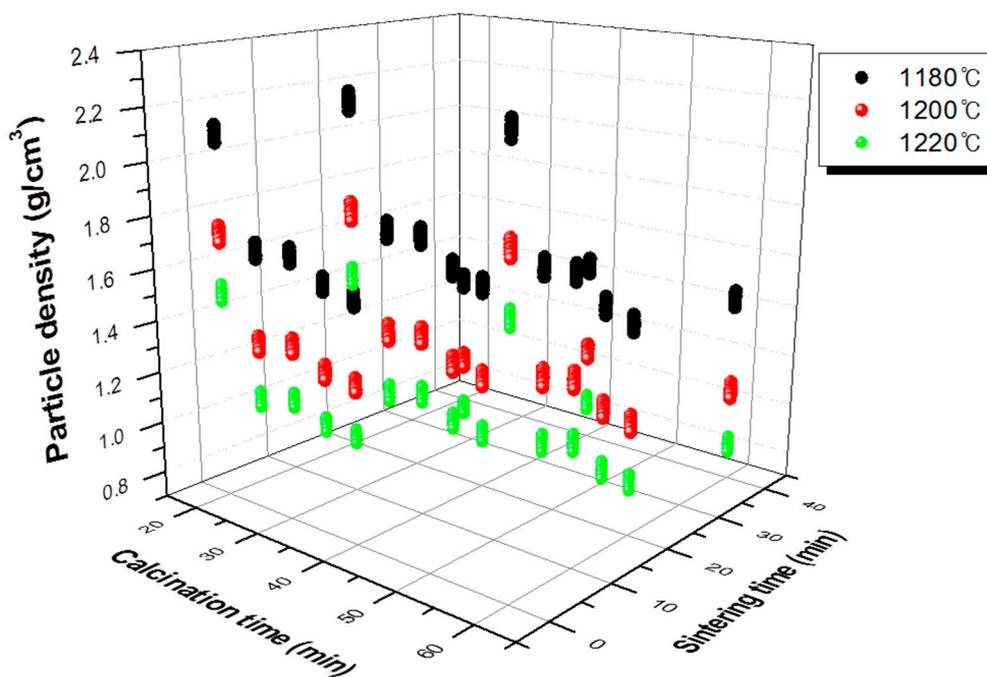


Figure 5. Expanded experiment results with the orthogonal array experiment design.

### 5.2. Evaluation of the Model

The results of the three machine-learning models are summarized in Table 3. As shown in Table 3, for  $R^2$ , SVR is the best model because this model is the largest and MSE, RMSE and MAE are the smallest. These values tend to improve with the number of data increases, given more data, this value will be improved. This means that a better model can be obtained by creating a model by expanding the experimental data than the modeling result using 18 data elements, which is the actual data. In fact, existing studies have reported that the creation of virtual data improves the  $R^2$  and MSE values of the model [47–49].

Table 3. Modeling performance.

	Linear Regression	Random Forest	SVR
$R^2$	0.799	0.783	0.933
MSE	0.029	0.031	0.009
RMSE	0.171	0.178	0.098
MAE	0.146	0.143	0.071

### 5.3. Experimental Evaluation of the Model Accuracy

Additional experiments were conducted to compare the actual experimental results and modeling results. This is shown in Table 4. The main parameters, sintering temperature and sintering time, were considered, and calcination time and drying time were fixed. In order to compare the actual experimental results with the predicted trends for each model, contour lines were based on the measured density values with the sintering temperature and time used as variables (Figure 6). As a result of this analysis, the tendency for the density to decrease as the sintering temperature increases and with longer sintering times is clearly visible. The predicted value of the single particle density differed according to the modeling method. For the value predicted through the orthogonal array experiment design, the maximum value tended to be rather large and the minimum value rather small. This likely occurs because not all interactions are reflected when predicting values using the orthogonal array experiment design. For this reason, while orthogonal array experimental design techniques are used in many fields [50–52], their experimental predictions are not well utilized. In a study by Chen [53], an optimal process was designed using an orthogonal array table, but a loss function value was used instead of an experimental value. In this prediction result, as described above, the result was somewhat overestimated in terms of the maximum and minimum values. The results predicted by the random forest method were suitably reflected in the trends of the sintering temperature and time but were predicted to be more biased toward the average than the actual values. On the other hand, the result predicted by SVR was somewhat inaccurate in terms of the maximum value part compared to the actual experiment, but the properties were predicted to be closest to the most commonly measured result among the three models. For a better comparison of the actual experiments with the prediction results, the results of measuring the particle density of aggregates sintered at temperatures of 1180, 1200, and 1220 °C are compared with the predicted values of each model, as shown in Figure 7. The results predicted by the orthogonal array experiment design are in good agreement with the actual measurement results. However, the single particle density was predicted to be much lower than the measured values at 20 min and 40 min at a temperature of 1220 °C. For the random forest method, the coincidence of the trend was confirmed, but the single particle density was predicted to be relatively high at 20 and 40 min, representing the time points at which the single particle density change is nearly saturated. In the SVR case, the single particle density value at 0 min was predicted to be lower, but it was confirmed that the remainder was predicted to be closer to the actual experimental value. In particular, under the conditions of 20 min and 40 min at 1220 °C, the closest accurate predicted value was obtained. In general, in statistics, various methods are tried and verified when analyzing data because it is impossible to know which method is better before analysis. In this study, SVR appeared as the most suitable modeling method, and the reason is thought to be due to the characteristics of individual methods. Linear regression requires strong assumptions: linear relationship between dependent variables and the target variable, normally distributed errors, homoscedasticity of errors and independence of the points. In general, it is very hard to satisfy the above assumptions in real world data. Decision tree ensembles including random forest tend to be overly complicated in the process of learning non-linear relationships with a small number of points. As a result, it is likely to be unstable in the phase of prediction of new data. In this study, we use the kernel SVR with a radial basis function. Unlike the above two models, it is known that the kernel SVR is good at learning complex and non-linear relationships between variables [54].

**Table 4.** Additional experimental conditions for model verification.

Process	Process Temp. (°C)	Variables
Drying and preheating	25~300 °C	60 min Fixed
	300~600 °C	60 min Fixed
Calcination	600 °C ~ B.S.A. temp.	20 min Fixed
Bloating Start and Activation (B.S.A.)	1180, 1190, 1200, 1210, 1220 °C	Time 0, 5, 10, 15, 20, 40 min

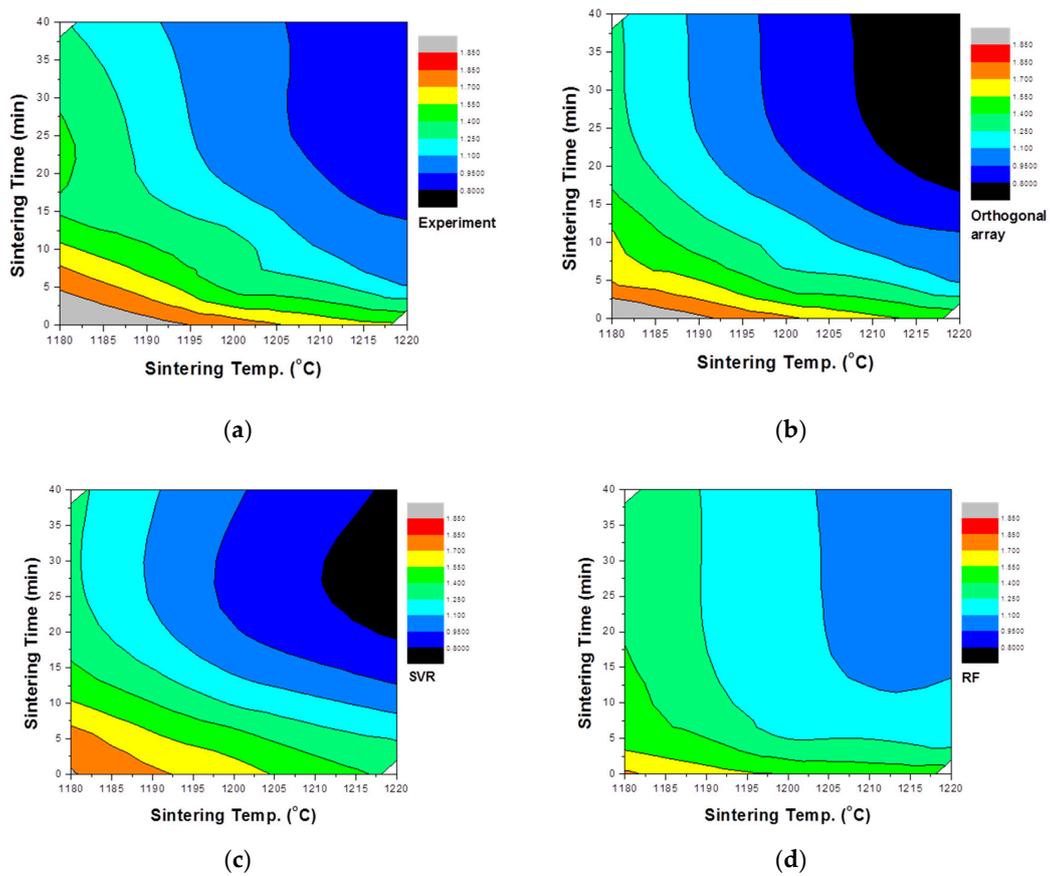


Figure 6. Particle density change of the aggregate with the soaking temperature and time as variables: (a) actual measurement, (b) orthogonal array-designed prediction, (c) SVR prediction, and (d) random forest prediction.

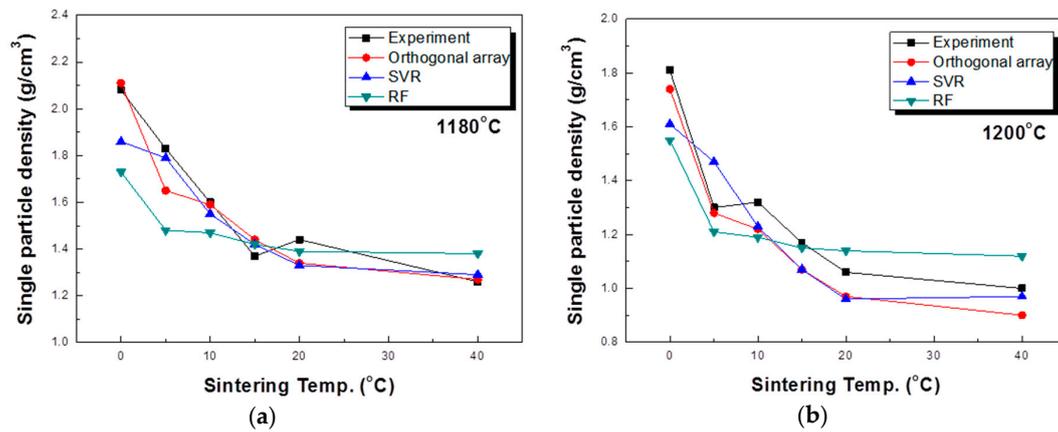
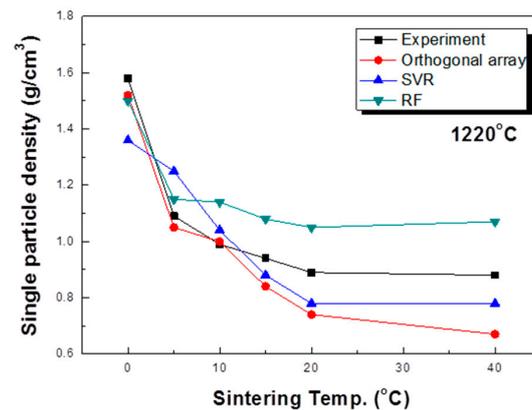


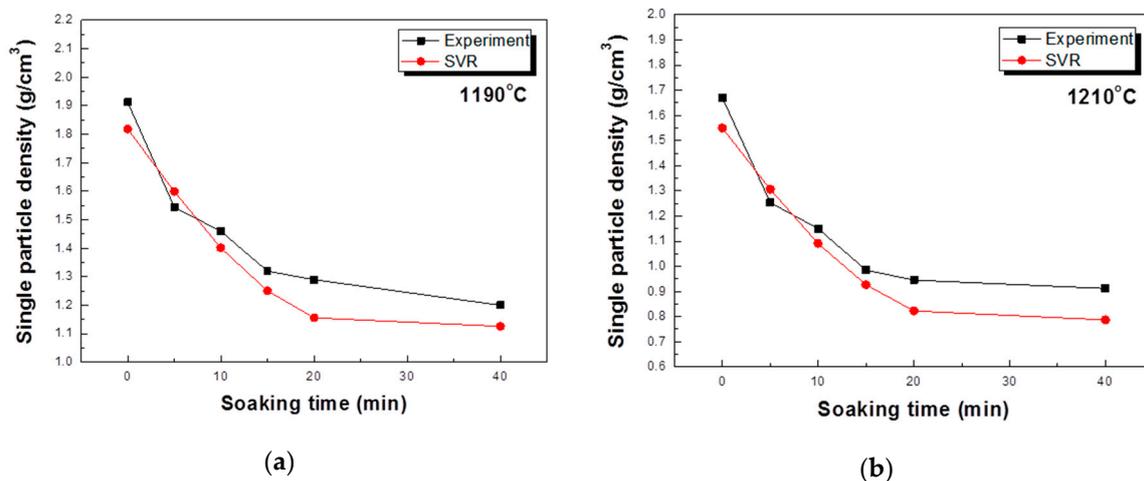
Figure 7. Cont.



(c)

**Figure 7.** Prediction of the single particle density according to each analysis method: (a) 1180 °C, (b) 1200 °C, and (c) 1220 °C.

In order to find out the predictive performance of the model for the unexperimented values, the physical properties were measured by firing the aggregate under conditions of 1190 °C and 1210 °C, and compared with the predicted values, and this is shown in Figure 8. At this time, the experimental conditions were fixed at r.t. ~300 °C and 300~600 °C for 60 min, and the condition at 600~bloating activation temperature was 20min. Most of the total predicted values were predicted close to the measured values. Experimental results with a sintering time of 0 to 15 min were quite consistent, and there was some error in the 20 to 40 min range, but the tendency to maintain the density was consistent. This means that modeling results using SVR can be predicted even in unexperimented areas, and the results are also suitable. In the expansion of data by an orthogonal sequence table, the results of the tested variables can be known, but the untested variables could not be defined.



**Figure 8.** Prediction of the single particle density according to the SVR method: (a) 1190 °C, (b) 1210 °C.

## 6. Conclusions

In this study, the experimental design of the change of particle density according to drying, calcination, and sintering conditions of artificial lightweight aggregate were performed using an orthogonal array table, and data were expanded. Modeling using machine-learning techniques were performed using the expanded data, and the following conclusions were drawn about the effectiveness of data expansion by an orthogonal array table and the appropriate modeling method:

1. The experimental design using the orthogonal array table and the expanded data gained through this method did not show a significant difference from the measured values.
2. The SVR model showed the best prediction accuracy among the reviewed models and generally showed good results even with untested variables.
3. Modeling by the random forest method predicts the trend of the process well. However, the result was predicted to be closer to the mean value than the actual value, and it could not be predicted for an untested part.
4. Through the experimental design and the expansion of the data by the orthogonal array table and modeling with the machine-learning technique, a model capable of efficiently predicting physical properties was realized.

In this experiment, only 18 data elements were actually tested for modeling, but these can be expanded to 486 through an orthogonal array table. This is a small amount of data in the field of data science but sufficient in the modeling field of scientific experiments. This means that the expansion of data through an orthogonal array table is useful for actual modeling. If there is a more considerable amount of data, the accuracy of the model can be improved. Hence, it is important to attempt this with more data. Orthogonal array experimental designs can be used to expand the number of variables, and in upcoming studies we will design experiments involving more process variables to evaluate the adequacy of the model. Also, studies on the correction of extended data and improvement of model reliability will be conducted.

**Author Contributions:** Conceptualization, K.G.L.; Methodology, Y.M.W.; Software, K.H.L. (Kang Hyuck Lee) and T.K.; Validation, T.K. and K.G.L.; Formal Analysis, Y.M.W.; Investigation, Y.M.W.; Resources, K.H.L. (Kang Hoon Lee); Data Curation, Y.M.W. and K.H.L. (Kang Hyuck Lee); Writing-Original Draft Preparation, Y.M.W.; Writing-Review and Editing, K.G.L. and K.H.L. (Kang Hoon Lee); Visualization, Y.M.W.; Supervision, K.G.L. and K.H.L. (Kang Hoon Lee); Project Administration, K.G.L.; Funding Acquisition, K.G.L. and K.H.L. (Kang Hoon Lee). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Basic Research Program through the National Research Foundation of Korea (NRF) grant number [2020R1F1A1067852] and the APC was funded by Korea Ministry of Environment “Prospective green technology innovation project [2020003160016]”.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Moreno-Maroto, J.M.; Uceda-Rodríguez, M.; Cobo-Ceacero, C.J.; Cotes-Palomino, T.; Martínez-García, C.; Alonso-Azcárate, J. Studying the feasibility of a selection of Southern European ceramic clays for the production of lightweight aggregates. *Constr. Build. Mater.* **2020**, *237*, 117583. [[CrossRef](#)]
2. Kim, D.S. Construction Material Demand Forecast in 2020. Korea Construction News. Available online: <http://www.conslove.co.kr/news/articleView.html?idxno=63080> (accessed on 28 October 2020).
3. Franus, M.; Panek, R.; Madej, J.; Franus, M. The properties of fly ash derived lightweight aggregates obtained using microwave radiation. *Constr. Build. Mater.* **2019**, *239*, 116677. [[CrossRef](#)]
4. Moreno-Maroto, J.M.; Uceda-Rodríguez, M.; Cobo-Ceacero, C.J.; de Hoces, M.C.; MartínLara, M.A.; Cotes-Palomino, T.; García, A.B.L.; Matínez-García, C. Recycling of ‘alperujo’ (olive pomace) as a key component in the sintering of lightweight aggregates. *J. Clean. Prod.* **2019**, *239*, 118041. [[CrossRef](#)]
5. Moreno-Maroto, J.M.; González-Corrochano, B.; Alonso-Azcárate, J.; García, C.M. A study on the valorization of a metallic ore mining tailing and its combination with polymeric wastes for lightweight aggregates production. *J. Clean. Prod.* **2019**, *212*, 997–1007. [[CrossRef](#)]
6. Guo, J.; Zhang, L.; Wang, W.; Zhang, H.; Liu, X.; Liu, M. Effect of SiO<sub>2</sub> and Al<sub>2</sub>O<sub>3</sub> on characteristics of lightweight aggregate made from sewage sludge and river sediment. *Ceram. Int.* **2017**, *44*, 4313–4319. [[CrossRef](#)]
7. Souza, M.M.; Anjos, M.A.S.; Sá, M.V.V.A.; Souza, N.S.L. Developing and classifying lightweight aggregates from sewage sludge and rice husk ash. *Case Stud. Constr. Mater.* **2020**, *12*, e00340. [[CrossRef](#)]
8. Chang, C.; Hong, G.; Lin, H. Artificial Lightweight Aggregate from Different Waste Materials. *Environ. Eng. Sci.* **2016**, *33*, 283–289. [[CrossRef](#)]

9. Wei, Y.; Cheng, S.; Ou, K.; Kuo, P.; Chung, T.; Xie, X. Effect of calcium compounds on lightweight aggregates prepared by firing a mixture of coal fly ash and waste glass. *Ceram. Int.* **2017**, *43*, 15573–15579. [[CrossRef](#)]
10. Wei, Y.; Cheng, S.; Chen, W.; Lu, Y.; Chen, K.; Wu, P. Influence of various sodium salt species on formation mechanism of lightweight aggregates made from coal fly ash-based material. *Constr. Build. Mater.* **2020**, *239*, 117890. [[CrossRef](#)]
11. Piszcz-Karaś, K.; Klein, M.; Hupka, J.; Łuczak, J. Utilization of shale cuttings in production of lightweight aggregates. *J. Environ. Manag.* **2019**, *231*, 232–240. [[CrossRef](#)]
12. Lee, K.G. Bloating Mechanism of Lightweight Aggregate with the Size. *J. Korean Ceram. Soc.* **2016**, *53*, 241–245. [[CrossRef](#)]
13. Kang, S.H.; Lee, K.G. Bloating Mechanism of Artificial Lightweight Aggregate for Recycling the Waste Glass. *J. Korean Ceram. Soc.* **2010**, *47*, 445–449. [[CrossRef](#)]
14. Li, Z.; Zhang, H.; Zhao, P.; He, X.; Duan, X. Manufacturing of Ultra-Light Ceramsite from Slate Wastes in Shangri-la, China. *J. Korean Ceram. Soc.* **2018**, *55*, 36–43. [[CrossRef](#)]
15. Kang, M.A.; Kang, S.G.; Lee, K.G.; Kim, Y.T. Fabrication of Artificial Light-Weight Aggregates of Uniform Bloating Properties Using a Temperature-Raising Sintering Method. *J. Korean Ceram. Soc.* **2012**, *49*, 161–166. [[CrossRef](#)]
16. Wie, Y.M.; Lee, K.G. Optimum Bloating-Activation Zone of Artificial Lightweight Aggregate by Dynamic Parameters. *Materials* **2019**, *12*, 267. [[CrossRef](#)]
17. Dondi, M.; Cappelletti, P.; D'Amore, M.; de Gennaro, R.; Graziano, S.F.; Langella, A.; Raimondo, M.; Zanelli, C. Lightweight aggregates from waste materials: Reappraisal of expansion behavior and prediction schemes for bloating. *Constr. Build. Mater.* **2016**, *127*, 394–409. [[CrossRef](#)]
18. Moreno-Maroto, J.M.; Cobo-Ceacero, C.J.; Uceda-Rodríguez, M.; Cotes-Palomino, T.; García, C.M.; Alonso-Azcárra, J. Unraveling the expansion mechanism in lightweight aggregates: Demonstrating that bloating barely requires gas. *Constr. Build. Mater.* **2020**, *247*, 118583. [[CrossRef](#)]
19. Wie, Y.M.; Lee, K.G.; Lee, K.H. Optimum conditions for unit processing of artificial lightweight aggregates using the Taguchi method. *J. Asian Ceram. Soc.* **2019**, *7*, 331–341. [[CrossRef](#)]
20. Gos, M.; Krzyszczyk, J.; Baranowski, P.; Murat, M. Combined TBATS and SVM model of minimum and maximum air temperatures applied to wheat yield prediction at different locations in Europe. *Agric. For. Meteorol.* **2020**, *281*, 107827. [[CrossRef](#)]
21. Ge, Y.; Zhao, S.; Zhao, X. A step-by-step classification algorithm of protein secondary structures based on double-layer SVM model. *Genomics* **2020**, *112*, 1941–1946. [[CrossRef](#)]
22. Gerhardt, N.; Schwolow, S.; Rohn, S.; Pérez-Cacho, P.R.; Galán-Soldevilla, H.; Arce, L.; Weller, P. Quality assessment of olive oils based on temperature-ramped HS-GC-IMS and sensory evaluation: Comparison of different processing approaches by LDA, kNN, and SVM. *Food Chem.* **2019**, *278*, 720–728. [[CrossRef](#)] [[PubMed](#)]
23. Li, M.; Wei, D.; Liu, T.; Liu, Y.; Yan, L.; Wei, Q.; Du, B.; Xu, W. EDTA functionalized magnetic biochar for Pb(II) removal: Adsorption performance, mechanism and SVM model prediction. *Sep. Purif. Technol.* **2019**, *227*, 115696. [[CrossRef](#)]
24. Fan, J.; Yue, W.; Wu, L.; Zhang, F.; Cai, H.; Wang, X.; Lu, X.; Xiang, Y. Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agric. For. Meteorol.* **2018**, *263*, 225–241. [[CrossRef](#)]
25. Yu, Y.; Zhang, C.; Gu, X.; Cui, Y. Expansion prediction of alkali aggregate reactivity-affected concrete structures using a hybrid soft computing method. *Neural Comput. Appl.* **2019**, *31*, 8641–8660. [[CrossRef](#)]
26. Sihag, P.; Kumar, M.; Singh, V. Enhanced soft computing for ensemble approach to estimate the compressive strength of high strength concrete. *J. Mater. Eng. Struct.* **2019**, *6*, 93–103.
27. Abd, A.M.; Abd, S.M. Modelling the strength of lightweight foamed concrete using support vector machine (SVM). *Case Stud. Constr. Mater.* **2017**, *6*, 8–15. [[CrossRef](#)]
28. Azimi-Pour, M.; Eskandari-Naddaf, H.; Pakzad, A. Linear and non-linear SVM prediction for fresh properties and compressive strength of high volume fly ash self-compacting concrete. *Constr. Build. Mater.* **2020**, *230*, 117021. [[CrossRef](#)]
29. Bolandi, H.; Banzhaf, W.; Lajnef, N.; Barri, K.; Alavi, A.H. An Intelligent Model for the Prediction of Bond Strength of FRP Bars in Concrete: A Soft Computing Approach. *Technologies* **2019**, *7*, 42. [[CrossRef](#)]

30. Solhmirzaei, R.; Salehi, H.; Kodur, V.; Naser, M.Z. Machine learning framework for predicting failure mode and shear capacity of ultra high performance concrete beams. *Eng. Struct.* **2020**, *224*, 111221. [[CrossRef](#)]
31. Korean Standards Association. *Testing Method for Density and Absorption of Coarse Aggregate*; KS F 2503:2007; Korean Standards Association: Seoul, Korea, 2007.
32. Sabarish, K.V.; Paul, P. An experimental analysis on structural beam with Taguchi orthogonal array. *Mater. Today Proc.* **2020**, *22*, 874–878. [[CrossRef](#)]
33. Zhang, F.; Wang, M.; Yang, M. Successful application of the Taguchi method to simulated soil erosion experiments at the slope scale under various conditions. *CATENA* **2021**, *196*, 104835. [[CrossRef](#)]
34. Prakash, K.S.; Gopal, P.M.; Karthik, S. Multi-objective optimization using Taguchi based grey relational analysis in turning of Rock dust reinforced Aluminum MMC. *Measurement* **2020**, *157*, 107664. [[CrossRef](#)]
35. Khalilarya, S.; Chitsaz, A.; Mojaver, P. Optimization of a combined heat and power system based gasification of municipal solid waste of Urmia University student dormitories via ANOVA and taguchi approaches. *Int. J. Hydrogen Energy* **2020**. [[CrossRef](#)]
36. Saarish, K.V.; Parvati, T.S. An anatomization of concrete elements with Taguchi optimization method. *Case Mater. Today Proc.* **2020**. [[CrossRef](#)]
37. Rathinam, N.; Dhinakaran, R.; Sharath, E. Optimizing process parameters to reduce blowholes in high pressure die casting using Taguchi methodology. *Mater. Today Proc.* **2020**. [[CrossRef](#)]
38. Mohd, U.; Pandulu, G.; Jayaseelan, R. Strength evaluation of eco-friendly concrete using Taguchi method. *Mater. Today Proc.* **2020**, *22*, 937–947. [[CrossRef](#)]
39. Sharifi, E.; Sadjadi, S.J.; Aliha, M.R.M.; Moniri, A. Optimization of high-strength self-consolidating concrete mix design using and improved Taguchi optimization method. *Constr. Build. Mater.* **2020**, *236*, 117547. [[CrossRef](#)]
40. Kechagias, J.D.; Aslani, K.-E.; Fountas, N.A.; Vaxevanidis, N.M.; Manolakos, D.E. A comparative investigation of Taguchi and full factorial design for machinability prediction in turning of titanium alloy. *Measurement* **2020**, *151*, 107213. [[CrossRef](#)]
41. Tyagi, R.; Shama, P.; Nautiyal, R.; Lakhera, A.K.; Kumar, V. Synthesis of quaternised guar gum using Taguchi L(16) orthogonal array. *Carbohydr. Polym.* **2020**, *237*, 116136. [[CrossRef](#)]
42. Feng, G.; Lei, S.; Guo, Y.; Shi, D.; Shen, J.B. Optimisation of air-distributor channel structural parameters based on Taguchi orthogonal design. *Case Stud. Therm. Eng.* **2020**, *21*, 100685. [[CrossRef](#)]
43. Wang, H.K.; Wang, Z.H.; Wang, M.C. Using the Taguchi method for optimization of the powder metallurgy forming process for Industry 3.5. *Comput. Ind. Eng.* **2020**, *148*, 106635. [[CrossRef](#)]
44. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
45. Drucker, H.; Burges, C.J.C.; Kaufman, L.; Smola, A.; Vapnik, V. Support vector regression machines. In Proceedings of the 9th International Conference on Neural Information Processing Systems, Denver, CO, USA, 3–5 December 1996; pp. 155–161.
46. Kōse, V.S.; Bayer, G. Bubble formation in the system waste glass-SiC and properties of such foam glasses. *Glastech. Ber.* **1982**, *55*, 151–160.
47. NezhadShokouhi, M.M.; Majidi, M.A.; Rasoolzadegan, A. Software defect prediction using over-sampling and feature extraction based on Mahalanobis distance. *J. Supercomput.* **2020**, *76*, 602–635. [[CrossRef](#)]
48. Zhu, T.; Lin, Y.; Liu, Y. Improving interpolation-based oversampling for imbalanced data. *Knowl. Based Syst.* **2020**, *187*, 104826. [[CrossRef](#)]
49. Liang, X.W.; Jiang, A.P.; Li, T.; Xue, Y.Y.; Wang, G.T. LR-SMOTE—An improved unbalanced data set oversampling based on K-means and SVM. *Knowl. Based Syst.* **2020**, *196*, 105845. [[CrossRef](#)]
50. Maazinejad, B.; Mohammadnia, O.; Ali, G.A.M.; Makhlof, A.S.H.; Nadagouda, M.N.; Sillanpää, M.; Asiri, A.M.; Agarwal, S.; Gupta, V.K.; Sadegh, H. Taguchi L<sub>9</sub>(3<sup>4</sup>) orthogonal array study based on methylene blue removal by single-walled carbon nanotubes-amine: Adsorption optimization using the experimental design method, kinetics, equilibrium and thermodynamic. *J. Mol. Liq.* **2020**, *298*, 112001. [[CrossRef](#)]
51. Akyalcin, S.; Akyalcin, L.; Bjørge, M. Optimization of desilication parameters of low-silica ZSM-12 by Taguchi method. *Microporous Mesoporous Mater.* **2019**, *273*, 256–264. [[CrossRef](#)]
52. Ayhan, V.; Cesur, Ç.Ç.I.; Çoban, A.; Ergen, G.; Çay, Y.; Kolip, A.; Özser, I. Optimization of the factors affecting performance and emissions in a diesel engine using biodiesel and EGR with Taguchi method. *Fuel* **2020**, *261*, 116371. [[CrossRef](#)]

53. Chen, H.-J.; Chang, S.-N.; Tang, C.-W. Application of the Taguchi Method for Optimizing the Process Parameters of Producing Lightweight Aggregates by Incorporating Tile Grinding Sludge with Reservoir Sediments. *Materials* **2017**, *10*, 1294. [[CrossRef](#)]
54. Burges, C.J. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167. [[CrossRef](#)]

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).