energies

MDPI

Article

Gradient Boosting Regression Tree Optimized with Slime Mould Algorithm to Predict the Higher Heating Value of Municipal Solid Waste

Esraa Q. Shehab [1], Farah Faaq Taha [1], Sabih Hashim Muhodir [2], Hamza Imran [3,*], Krzysztof Adam Ostrowski [4,*] and Marcin Piechaczek [4]

[1] Department of Civil Engineering, College of Engineering, University of Diyala, Baqubah 32001, Iraq; esraa_qasim@uodiyala.edu.iq (E.Q.S.); con.mango.msc.farah@uodiyala.edu.iq (F.F.T.)
[2] Department of Architectural Engineering, Cihan University Erbil, Erbil 44001, Iraq; sabih.alzuhairy@cihanuniversity.edu.iq
[3] Department of Construction and Project, Al-Karkh University of Science, Baghdad 10081, Iraq
[4] Faculty of Civil Engineering, Cracow University of Technology, Warszawska 24, 31-155 Cracow, Poland; marcin.piechaczek1@doktorant.pk.edu.pl
* Correspondence: hamza.ali1990@kus.edu.iq (H.I.); krzysztof.ostrowski.1@pk.edu.pl (K.A.O.)

Abstract: The production of municipal solid waste (MSW) has led to an unprecedented level of environmental pollution, worsening the global challenges posed by climate change. Researchers and policymakers have recently made significant strides in the field of sustainable and renewable energy sources, which are viable from technological, environmental, and economic perspectives. Consequently, the waste-to-energy programs enhance nations' socioeconomic status while positively impacting the environment. To predict the higher heating value (HHV) of MSW fuel based on carbon, hydrogen, oxygen, nitrogen, and sulfur content, the current study introduces a Gradient Boosting Regression Tree (GBRT) model optimized with the Slime Mold Algorithm (SMA). This model was evaluated using an additional 50 data points after being trained with 202 MSW biomass data points. The performance of the model was assessed using three metrics: root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2). The results indicated that our model outperformed previously developed models in terms of accuracy and reliability. Additionally, a graphical user interface (GUI) was developed to facilitate the practical application of the model, allowing users to easily input data and receive predictions on the enthalpy of the combustion of MSW fuel.

Keywords: Gradient Boosting Regression Tree; machine learning; municipal solid waste; higher heating value

Citation: Shehab, E.Q.; Taha, F.F.; Muhodir, S.H.; Imran, H.; Ostrowski, K.A.; Piechaczek, M. Gradient Boosting Regression Tree Optimized with Slime Mould Algorithm to Predict the Higher Heating Value of Municipal Solid Waste. *Energies* **2024**, *17*, 4213. https://doi.org/10.3390/en17174213

Academic Editor: Hamed Aly

Received: 25 July 2024
Revised: 13 August 2024
Accepted: 19 August 2024
Published: 23 August 2024

1. Introduction

One of the essential elements of sustainable development is the availability of energy resources [1]. Beyond human labor, sufficient energy is the primary economic driver in industrial societies, as it is indispensable for social welfare, economic growth, improving living standards, and ensuring societal security [2]. The concept of sustainable energy will be realized if energy is produced and consumed in a manner that supports long-term human progress across all domains—economic, social, and environmental. Therefore, sustainable development necessitates a reliable and sustainable energy supply [3].

Over the past decade, fossil fuels have increasingly become the preferred energy source, overshadowing other alternatives [4]. Growing concerns about the threat of global warming and the gradual depletion of traditional fossil fuels have heightened the urgency to develop renewable and sustainable energy sources [5]. Additionally, renewing these resources is impractical due to the millions of years needed for the formation of hydrocarbon chains. In terms of climate change, the use of fossil fuels significantly contributes to the

emission of harmful pollutants such as CO, CH_4, H_2S, and N_2O. Among these, CH_4 has the most substantial impact on climate change, while NO_x emissions pose serious health risks, including respiratory issues [6]. Due to these factors, both industrialized and emerging nations have been increasingly focusing on renewable energy sources such as solar, wind, geothermal, and biomass. Their efforts aim to diversify energy sources, reduce reliance on traditional energy carriers, and address environmental concerns [7].

In 2016, urban areas worldwide produced 2.01 billion tons of solid waste, equating to 0.74 kg per person per day. This annual waste generation is projected to increase to 3.4 billion tons by 2050, representing a 70% rise from 2016, driven by the current rate of urbanization and population growth [8,9]. Waste management expenses account for 20–50% of municipal budgets [8]. In underdeveloped nations, nearly 90% of waste is often disposed of improperly or incinerated in the open, exacerbating the negative impacts of waste management [8].

Effective waste management is an integral element of the European Union's policy, in accordance with Directive 2018/850 [10]. This directive aims to reduce the amount of municipal waste sent to landfill to 10% of the total mass of generated waste by the year 2035. Currently, this level stands at approximately 20%. Successful implementation of these objectives requires innovative solutions and continuous monitoring of progress in the field of waste management, which is crucial for achieving sustainable development and environmental protection.

The adoption of the waste-to-energy concept is driven by the need to combat the aforementioned hazard, as it reduces the landfilling of municipal solid waste and decreases the reliance on traditional hydrocarbon fuels [11]. Waste-derived energy is increasingly being incorporated into our energy feedstock.

Waste-to-energy conversion can be achieved through various methods, such as gasification, fermentation and distillation, anaerobic digestion, or waste incineration [12]. Among these, waste incineration is one of the most frequently chosen and modernized solutions for converting waste into energy (Figure 1). The practice of incinerating waste is a technique known for generations, raising questions about whether this process aligns with the trend of sustainable waste management [13–15]. To improve the environmental sustainability of the process, waste incineration has been modernized through the use of filters and modifications to combustion temperatures. These enhancements aim to reduce the emission of harmful substances and improve the energy efficiency of the process.

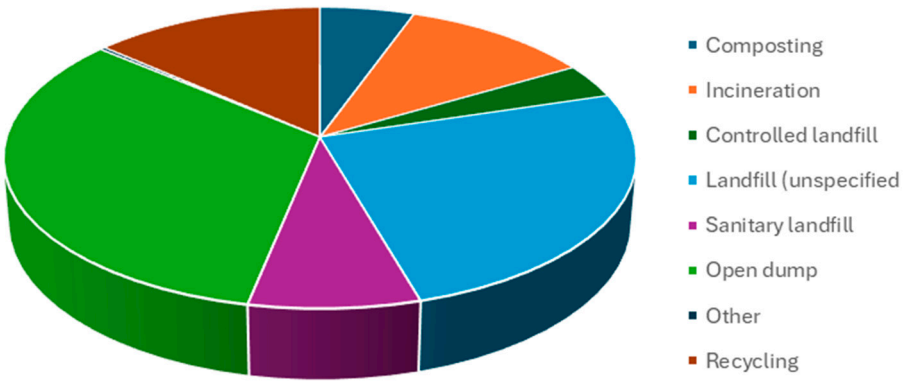

Figure 1. Global treatment and disposal of waste (percent) [16].

Therefore, calculating the combustion enthalpy from common components such as carbon, moisture, oxygen, nitrogen, sulfur, and ash enables engineers and policymakers to make informed decisions swiftly. Additionally, a thorough understanding of the key characteristics of municipal solid waste (MSW), including its elemental composition and combustion enthalpy, is essential for developing modern systems that can convert waste into various energy resources, such as electricity, heat, and transportation fuels. Enthalpy of combustion, which is also known as higher heating value (HHV), is a critical parameter that

determines the energy content of MSW. In contrast, the lower heating value (LHV) measures the energy released when the same quantity of fuel is fully burned but the vapor formed is allowed to remain in its gaseous state [17]. A bomb calorimeter is a labor- and time-intensive tool used in the traditional method of determining the HHV of fuel samples. Many studies use proximate and ultimate analysis of fuel samples simultaneously to determine the HHV. Compared to bomb calorimeters and proximate analysis, ultimate analysis offers several advantages. It provides a detailed description of each fuel's components in terms of carbon (C), nitrogen (N), oxygen (O), sulfur (S), and hydrogen (H). Furthermore, using the percentages of C, H, O, N, and S helps to improve the fuel's combustion process and predicts the heating values more accurately than proximate analysis [18]. Thus, it will be especially beneficial to quickly and affordably develop a tool that can reliably forecast the HHVs of various wastes using data mining techniques and historical data.

Recent times have seen the application of artificial intelligence as a data-driven methodology to improve the efficiency of waste and biofuel systems [19].

Through the keyword set "waste incineration machine learning", the Scopus library catalog was analyzed, yielding 115 publications from the past 32 years. Notably, 96% of these publications were published within the last 5 years. Scopus was chosen due to its extensive number of publications containing the relevant keywords compared to other databases.

Subsequently, a visualization illustrating the frequency of occurrence and the relationships between relevant keywords in the Scopus database was created using VOSviewer software (version 1.6.17) (Figure 2). To present these results in a readable manner, the minimum number of connections between keywords was set to three, reducing the overall number of keywords from 1770 to 204. Analyzing the obtained visualization, strong connections with the fields of materials science, artificial intelligence (AI), and environmental protection were observed.

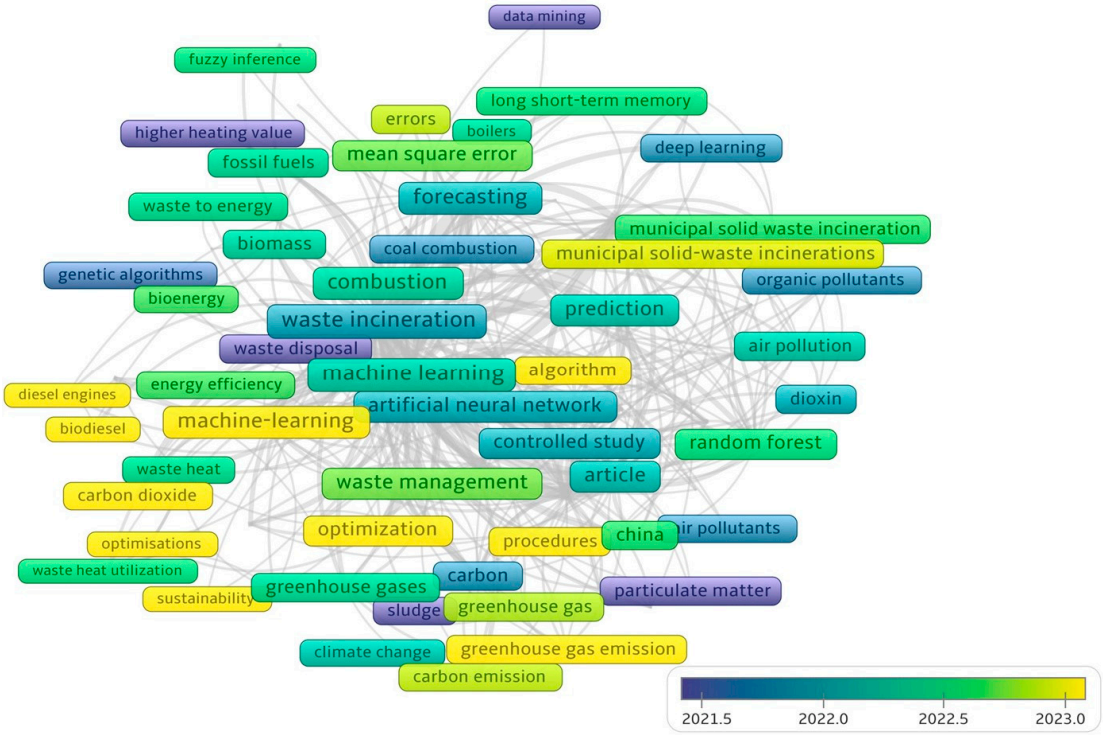

Figure 2. Visualization of the incidences of individual keywords and their relationships (accessed on 20 June 2024).

The conducted analysis indicates the importance of a detailed examination of the impact of using machine learning for municipal waste incineration.

According to Naveed et al. [20], this approach complies with social, environmental, and governance (ESG) standards, skillfully addressing environmental problems and encouraging sustainable business practices. Several studies in the literature provide methods for estimating HHVs from proximate analysis (P), ultimate analysis (U), or both (UP) datasets. For example, one of the studies used four machine learning methods (Radial Basis Function Artificial Neural Network (RBF-ANN), Multi-Layer Perceptron Artificial Neural Network (MLP-ANN), Support Vector Machine (SVM), and Adaptive Neuro-Fuzzy Inference System (ANFIS) to predict MSW HHVs using carbon, water, hydrogen, oxygen, nitrogen, sulfur, and ash data [21]. The RBF-ANN model showed the best accuracy with a 0.45% MAPE. Another study developed ANN and Particle Swarm Optimization (PSO) models to predict MSW HHVs using proximate and ultimate analysis [22]. The ANN-4 model, combining both analyses, showed the best accuracy, with deviations under 10%. Linear models were unsuitable, but PSO models improved prediction performance significantly. Also, the study by [23] used robust estimators to address non-normality and multicollinearity in regression models. The best model, using the robust K-L estimator, showed high accuracy, with an adjusted R^2 of 0.9710 and minimal errors. Furthermore, the study by [24] presented a rapid, cost-effective method using machine learning and ultimate analysis to estimate HHVs for diverse municipal solid wastes. Gene Expression Programming (GEP), SVM, and Feedforward Neural Network (FFNN) models were developed, showing high accuracy (R^2 up to 0.978) and low numbers of errors (AAE 0.87–1.12%). Last but not least, the study by [25] introduced an ANFIS model optimized with PSO to predict MSW fuel enthalpy based on moisture, carbon, hydrogen, oxygen, nitrogen, sulfur, and ash content. The model showed superior accuracy compared to existing models.

Drawing from earlier studies, adjusting a machine learning model's hyperparameters with optimization algorithms is a highly effective technique for enhancing prediction accuracy. This method, referred to as a hybrid model, integrates various algorithms to achieve superior results. The Sparrow Search Algorithm (SSA), the Butterfly Optimization Algorithm (BOA), the Particle Swarm Optimization algorithm (PSO), the Whale Optimization Algorithm (WOA), and others are examples of optimization algorithms. Many hybrid models have been developed and used in HHV energy domains in recent years. For instance, the paper by [26] presented metaheuristic-based ANN models optimized with PSO (ANN-PSO) and multilinear regression models to predict the higher heating values (HHVs) of solid fuels. The ANN-PSO models demonstrated superior performance, evidenced by better statistical parameters and predictive ability compared to multilinear regression models. Another study introduced a particle swarm optimization support vector regression model for the accurate gross calorific value prediction of coal, using radial basis, linear, and polynomial kernel functions [27]. Compared to classification and regression trees, multiple linear regression, and principal component analysis models, the Particle Swarm Optimization Support Vector Regression model with a radial basis function demonstrated superior performance, showing high efficiency and accuracy in predicting gross calorific values. Furthermore, the study by [28] optimized the ANFIS model with Particle Swarm Optimization (PSO) and genetic algorithms (GAs) to predict the lower heating value (LHV) of waste. The grid partitioning-clustered PSO-ANFIS model with a triangular input membership function showed superior accuracy, demonstrating the robustness of evolutionary-based neuro-fuzzy models. Moreover, the study by [29] enhanced Random Forest Regression (RFR) for predicting hydrogen and nitrogen in gasification processes using Snake Optimization (SO) and an Equilibrium Optimizer (EO) model. The optimized models RFEO and RFSO achieved superior performance, with R^2 values of 0.997 and better accuracy metrics than the baseline RFR model.

Even though numerous machine learning studies on predicting the HHV of MSW have been published, most relied on smaller datasets and lacked an interpretable ML model. Moreover, no research has used the Gradient Boosting Regression Tree (GBRT) to estimate the HHV of MSW until now. Friedman [30] presented the GBRT, an innovative and effective tree-based machine learning model. By automating parallel computing, the GBRT

model can save computing costs and prevent overfitting during training. The industrial community and Kaggle competitions have extensively recognized the performance of the GBRT model [31]. In addition to examining the utility of the GBRT model, this study focuses on its optimization, which is crucial for enhancing ML model performance and reducing computing costs.

2. Research Significance

The aim of the presented research was to apply the Slime Mould Algorithm (SMA) with the Gradient Boosting Regression Tree (GBRT) model to predict the higher heating value (HHV) of municipal solid waste. The SMA was used to select the best hyperparameters for the GBRT model, thereby enhancing its predictive power.

Several performance assessment metrics were computed, including the correlation coefficient (R), determination coefficient (R^2), root mean square error (RMSE), and mean absolute error (MAE), to evaluate the model's effectiveness.

Additionally, a comparative analysis was conducted between the SMA-GBRT model and previously developed models.

3. Materials and Methods

The goal of our work is to forecast the HHV of MSW by combining the GBRT and the SMA into a hybrid model called the GBRT-SMA. In this hybrid model, the GBRT serves as the primary prediction algorithm, creating a function that can be used to calculate the HHV of MSW based on a collection of explanatory factors such as carbon, hydrogen, oxygen, nitrogen, and sulfur. Despite the robustness of the GBRT algorithm, its effectiveness heavily depends on the selection of its parameters. Consequently, the SMA is employed to determine the optimal combination of GBRT settings. Following the determination of these ideal parameters, the associated GBRT model is trained to provide the final forecast. Figure 3 depicts the complete workflow of the SMA-GBRT model, which comprises three stages: data preprocessing, hyperparameter tuning, and final prediction. During the preprocessing stage, the input data is standardized using the Z-score formula to achieve a mean of zero and approximately comparable magnitudes. The formula is as follows:

$$X_{\mathrm{N}} = \frac{X_{\mathrm{O}} - m_{\mathrm{X}}}{s_{\mathrm{X}}} \tag{1}$$

where m_{X} and s_{X} represent the mean and standard deviation of the feature under consideration across the entirety of the input data, and X_{N} and X_{O} are the normalized and original feature variables, respectively. The entire data set is then divided into training and testing subsets. The model is trained and the parameters are optimized using the training data set. To enable the SMA to select the optimal GBRT parameters during the parameter optimization stage, we must provide a mechanism to evaluate the quality of a given set of parameters. To achieve this, we develop an objective/cost function based on k-fold cross-validation and mean square error (MSE). The training data set is further divided into k, nearly equal-sized subsets or folds. For a given set of parameters, the corresponding GBRT model is trained and validated multiple times using various combinations of training and validation methods. For each fold, the model is validated against the data in the current fold after being trained using data from the previous $k-1$ folds. The average of the validation MSE is the definition of the cost function. To be more precise

$$\mathrm{CF} = \frac{1}{k}\sum_{i=1}^{k} \mathrm{MSE}_i \tag{2}$$

where, in the case when the training set is the union of multiple folds, MSE_i represents the validation MSE associated with the i_{th} fold:

$$\frac{\sum_{j \in S_i} \left(Y_{A,j} - Y_{P,j}\right)^2}{|S_i|} \quad (3)$$

where the actual output for the jth data set is $Y_{A,j}$, the corresponding projected output is $Y_{P,j}$, the index set of the ith fold is S_i, and $|.|$ indicates the cardinal of a set.

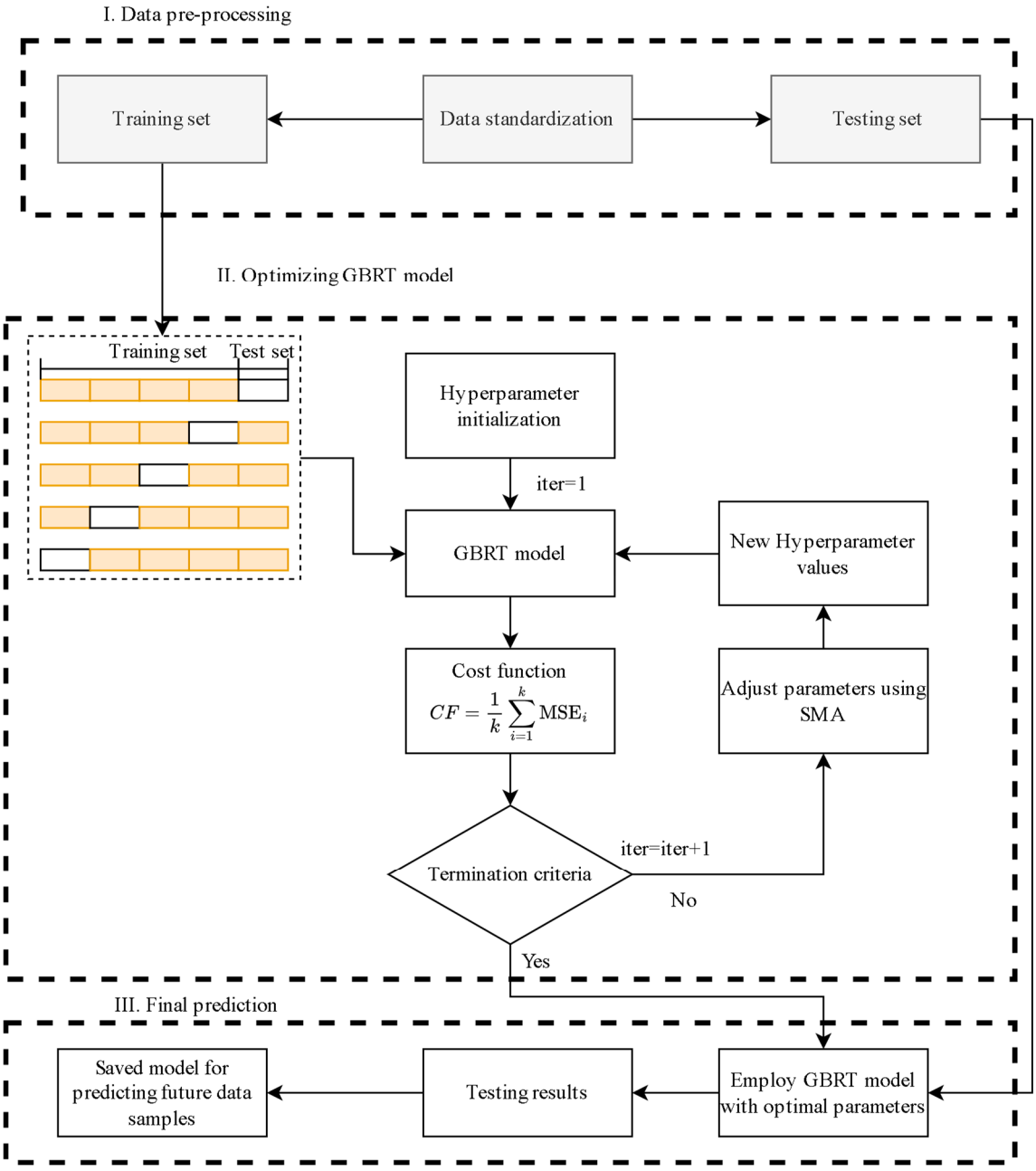

Figure 3. Research methodology.

When a predetermined maximum number of iterations has been reached or the cost function does not improve (decrease) after a specified number of iterations, the SMA search is terminated. In the final prediction stage, the corresponding GBRT model is employed

once the optimal set of parameters has been established. At this point, the estimated HHVs can be obtained and recorded.

The machine learning (ML) models in this work were implemented and assessed using the Python programming language, specifically through Google Colab and its related libraries, such as Scikit-learn (version 1.4.2). The sklearn package offers many effective techniques for statistical modeling and machine learning, including dimensionality reduction, clustering, regression, and classification. The primary use of open-source Python (3.7) modules and tools in this study resulted in minimal direct software costs. Notably, our methodology proved to be more affordable than conventional experimental approaches for measuring higher heating values (HHVs), due to the utilization of open-source software and publicly accessible computing resources.

3.1. An Overview of Modeling Methods

3.1.1. Classification and Regression Tree (CART)

The CART evolves into a binary decision tree by providing a "YES or NO" response to a binary classification of the input feature space. This model is widely used because it can help to solve problems in both regression and classification by discretizing continuous feature variables [32].

A regression tree recursively builds a multinomial tree to divide a new input feature space area *Rm* and its corresponding node output value *cm*. It splits the input features in the dataset based on feature attributes and feature thresholds, using the sum of squared errors as the evaluation index for the segmentation effect of the input feature space area. This process continues until the set generation conditions are satisfied [33]. The effectiveness of regional segmentation increases as the sum of squared errors decreases. Consequently, the splitting node selection process and the spatial and regional division of the input features significantly impact the prediction accuracy of the regression trees.

The input space $R = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)\}$ should be set up, with x_i representing the input characteristics, each being a multidimensional vector, $x_i \in \mathbb{R}_+^d$, and y_i representing the output, which is a positive scalar, $y_i \in \mathbb{R}_+$. After evaluating each splitting feature (j) and splitting node (s), calculate the minimal squared error, choose the best pair (j,s) for dividing the input feature's space and region, and produce two mutually disjointed sub-regions:

$$\min_{j,s}\left[\min_{c_1}\sum_{x_i \in R_1(j,s)}(y_i - c_1)^2 + \min_{c_2}\sum_{x_i \in R_2(j,s)}(y_i - c_2)^2\right] \quad (4)$$

$$R_1(j,s) = \left\{x \mid x^{(j)} \leqslant s\right\}, R_2(j,s) = \left\{x \mid x^{(j)} > s\right\} \quad (5)$$

Divide R into m separated segments, i.e., R_1, R_2, R_3, …, R_m, using the best pair (j, s) for space and region division of the input feature. Each R_m represents a specific region in the input feature space after several splits. The output values c_m for each region R_m are then averaged:

$$c_m = \text{ave}(y_i \mid x_i \in R_m) \quad (6)$$

$$f(x) = \sum_{m=1}^{M} c_m I(x \in R_m), I = \left\{\begin{array}{ll} 1, & x \in R_m \\ 0, & x \notin R_m \end{array}\right\} \quad (7)$$

To provide as many accurate prediction results as possible, the regression tree model divides the feature space and area during the dataset's training phase, expanding the tree model's branches. This process can lead to overfitting by adopting certain features from the training samples as universal attributes shared by all samples. Furthermore, instead of utilizing a feedback iteration process, a single regression tree creates tree nodes through a single feature selection, increasing the likelihood that the final output will yield a partially optimal solution.

3.1.2. Gradient Boosting Regression Tree (GBRT)

The GBRT, an ensemble learning method built on the boosting framework, was developed by Friedman [30] to address the drawbacks of the regression tree model mentioned above. As shown in Figure 4, the GBRT generates a new weak learner in the direction of the gradient of residual reduction using multiple regression trees as weak learners and the Forward Stagewise method to lower the residuals of the previous iteration's weak learner. Meanwhile, it adjusts the sample data weights to generate new weak learners through iterations, enabling the full learning of sample data features and enhancing the weak learners' prediction accuracy [34]. To produce prediction results with greater accuracy, the outcomes of all weak learners are ultimately combined linearly using an additive model.

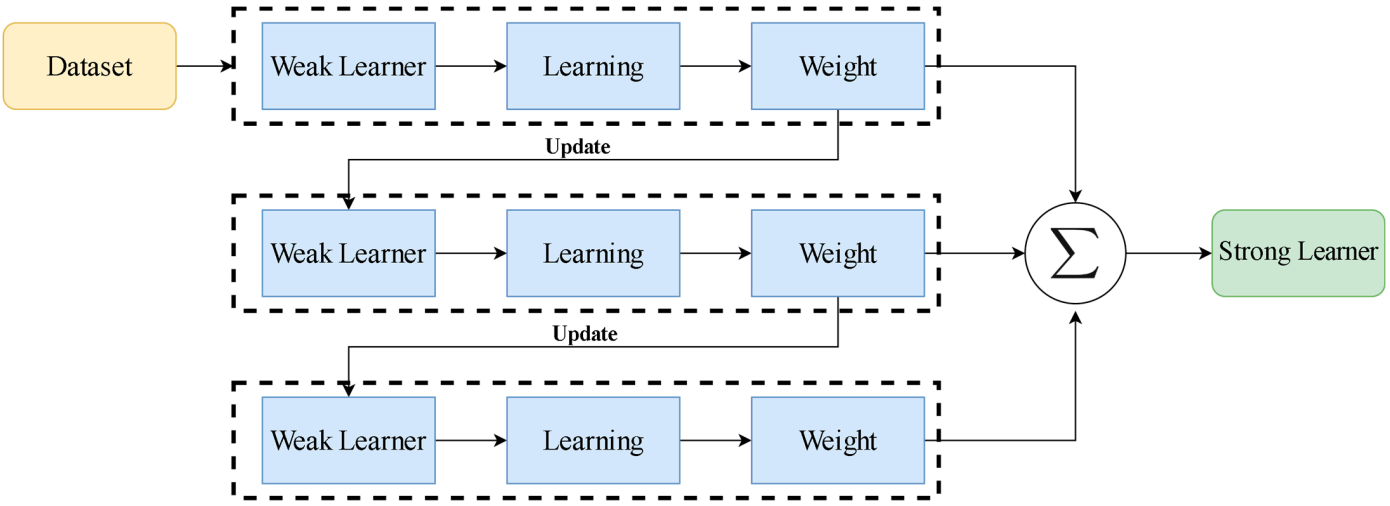

Figure 4. Workflow of Gradient Boosting Regression Trees (GBRTs).

Assign the following values to the input training data: $T = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)\}$ where y_i the output sample and x_i denotes the input sample. The square error function is used as the loss function $L(y, f(x))$, which is defined as follows:

$$L(y, f(x)) = \left[y - f(x)\right]^2 \tag{8}$$

The formulation for the GBRT's weak learners is the same as expression (8) above, as the GBRT employs multiple regression trees as weak learners. The following outlines the precise stages of the GBRT algorithm:

(a) Start with weak learners:

$$f_0(x) = \underset{c}{argmin} \sum_{i=1}^{N} L(y_i, c) \tag{9}$$

(b) Create m weak learners:

Determine the residual $(r_{m,i})$ of each sample $(x_i, y_i), i = 1, 2, 3, \ldots, n$ which is the loss function's negative gradient.

$$r_{m,i} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x)}\right]_{f(x)=f_{m-1}(x)}, i = 1, \ldots, N \tag{10}$$

Equation (10) describes the process of calculating the residuals for each sample in the training dataset at each iteration. These residuals represent the differences between the actual target values and the predicted values from the model. In the context of the GBRT, the residuals are computed as the negative gradient of the loss function with respect to the

current model's predictions. Weak learners are trained using the residuals and, ultimately, m independent regression trees $f_1(x)$, $f_2(x)$, …, $f_m(x)$, m = 1, 2, …, M are formed via iteration. Assume that each regression tree has j leaf nodes (where j = 1, 2, …, J) and that these leaf nodes split the input space into j mutually disjoint subregions, R_{m1}, R_{m2}, …, R_{mj}. Next, determine the best forecast value $c_{m,j}$ inside the R_{mj} area. The loss function's smallest value in the mj area is denoted by $c_{m,j}$.

$$c_{m,j} = \underset{c}{argmin} \sum_{x_i \in R_{m,j}} L(y_i, F_{m-1}(x) + c) \tag{11}$$

(c) Update the weak learner $f_m(x)$:

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J_m} c_{m,j} I\left(x \in R_{m,j}\right) \tag{12}$$

(d) The final gradient lifting regression tree expression is achieved following M rounds of iterations

$$f(x) = f_0(x) + \sum_{m=1}^{M} \sum_{j=1}^{J} c_{mj} I\left(x \in R_{mj}\right) \tag{13}$$

3.2. Slime Mould Algorithm (SMA)

The intelligent behavior of a mould known as Slime Mould served as the model for the SMA. Additionally, the mould exhibits intelligent behavior and has highly quick, error-free routing. Multicellular and reproductive structures are created when single-celled organisms of this type of mould come together. They monitor food quality and risk, and they meticulously balance their diets. The mould's starting position will be determined by [35], as follows:

$$x_{i,k} = lb + \text{rand}\ (0,1) \times (ub - lb) \tag{14}$$

where $x_{i,k}$ is the mould's initial position, and lb and ub are the lower and upper bounds chosen for each solution or attribute, respectively. Slime Mould approaches and routes its prey by releasing an odorous trail of food into the atmosphere. Based on the bait odor, the routing behavior of mucosal mould is represented by the following equation [35]:

$$\overrightarrow{X(t+1)} = \begin{cases} \overrightarrow{X_b(t)} + \overrightarrow{vb} \cdot \left(\overrightarrow{\dot{W}} \cdot \overrightarrow{X_A(t)} - \overrightarrow{X_B(t)} \right) r < p \\ \overrightarrow{vc} \cdot \overrightarrow{X(t)} \, r \geq p \end{cases} \tag{15}$$

When the current iteration is denoted by t, the parameter $\overrightarrow{vc}$ decreases linearly from 1 to 0, and $\overrightarrow{vb}$ is a variable within the range [−a, a]. The location of the Slime Mould with the highest concentration of odor at iteration t is indicated by $\overrightarrow{X_B(t)}$. The position of the Slime Mould at iteration t is indicated by $\overrightarrow{X(t)}$, and the position at the next iteration is indicated by $\overrightarrow{X(t+1)}$. Two randomly selected positions of the Slime Mould are denoted by $\overrightarrow{X_A(t)}$ and $\overrightarrow{X_b(t)}$, and the weight of the Slime Mould is represented by $\overrightarrow{\dot{W}}$. The value of p is obtained from Kumar et al. [36]:

$$p = \tanh |S(i) - DF| \tag{16}$$

When the number of cells in the Slime Mould is indicated by $i = 1, 2, \ldots, N$ and N, the quantity of fitness is shown by $S(i)$, and the greatest fitness attained across all iterations is represented by DF. The variable $\overrightarrow{vb}$ is obtained from the following equation:

$$\overrightarrow{vb} = [-a \cdot a] \tag{17}$$

$$a = \operatorname{arctanh}\left(-\left(\frac{t}{max_t}\right) + 1\right)$$

The following equation can also be used to calculate $\vec{W}$:

$$\overrightarrow{W(\text{smell index}\ (l))} = \begin{cases} 1 + r \cdot \log\left(\frac{bF - S(i)}{bF - wF} + 1\right), \text{ condition} \\ 1 - r \cdot \log\left(\frac{bF - S(i)}{bF - wF} + 1\right), \text{ others} \end{cases} \tag{18}$$

$$\text{smell index} = \text{sort}\ (S) \tag{19}$$

In the first half of the population, that condition equals $S(i)$. A random integer in the interval [0, 1] is represented by U, the best fitness is bF, the worst fitness is wF, and the sequence of fitness values is $SmellInex$. Additionally, the following formula is used to update the positions of the Slime Moulds [35]:

$$\overrightarrow{X^*} = \begin{cases} \text{rand}\ (UB - LB) + LB\ \text{rand} & < z \\ \overrightarrow{X_b(t)} + \overrightarrow{vb} \cdot \left(\vec{W} \cdot \overrightarrow{X_A(t)} - \overrightarrow{X_B(t)}\right) & r < p \\ \overrightarrow{vc} \cdot \overrightarrow{X(t)} & r \geq p \end{cases} \tag{20}$$

In the parameter setup experiment, the $z$value will be explored. Here, rand and r are random parameters within the interval [0, 1], and LB and UB are the lower and upper limits of the search interval, respectively. This process is repeated until the stop condition is satisfied. The location of the optimal characteristics is then indicated by the output $\overrightarrow{X^*}$ [35].

4. Database Used

Bagheri et al. [24] provided a dataset that included 252 distinct waste samples of municipal origin that had been experimentally examined. These samples represented the principal categories of municipal solid waste involved: twelve rubber and leather wastes, twenty-nine MSW mixtures (hereafter referred to as MSWs), thirty-four plastic wastes, thirty paper wastes, fifty-three other wastes, twelve textile wastes, sixty-one wood wastes, and twenty types of sewage sludge. The primary MSW components, such as C, O, H, N, and S, were the input variables, while the HHV (MJ/kg) was the output parameter. Following a randomization process, the 252 datasets were split into two groups, with 80% of the data being utilized for training and 20% for testing the models.

As seen in Figure 5, the correlation between two parameters can be measured using Pearson correlation coefficients (r) Higher r values indicate a stronger correlation between the two factors. In the MSW database, the independent variables (C, O, H, N, and S) and the dependent variable (HHV) have r values of 0.95, −0.49, 0.89, −0.21, and 0.057, respectively. This means that the HHV is primarily correlated with C and H. Conversely, there are weak and negative correlations between the values of r for O, N, S, and HHV. Based on scientific study of the properties of biomass energy and empirical findings, these relationships are consistent with the basic concept of biomass combustion [37,38]. Data preparation is essential for creating machine learning models as it ensures that the model is not trained with biased data. Consequently, the data are transformed using the Z-score normalization approach before being fed into the machine learning algorithms. Table 1

displays the coverage of the data points and the statistical analysis of the final elements and the HHV.

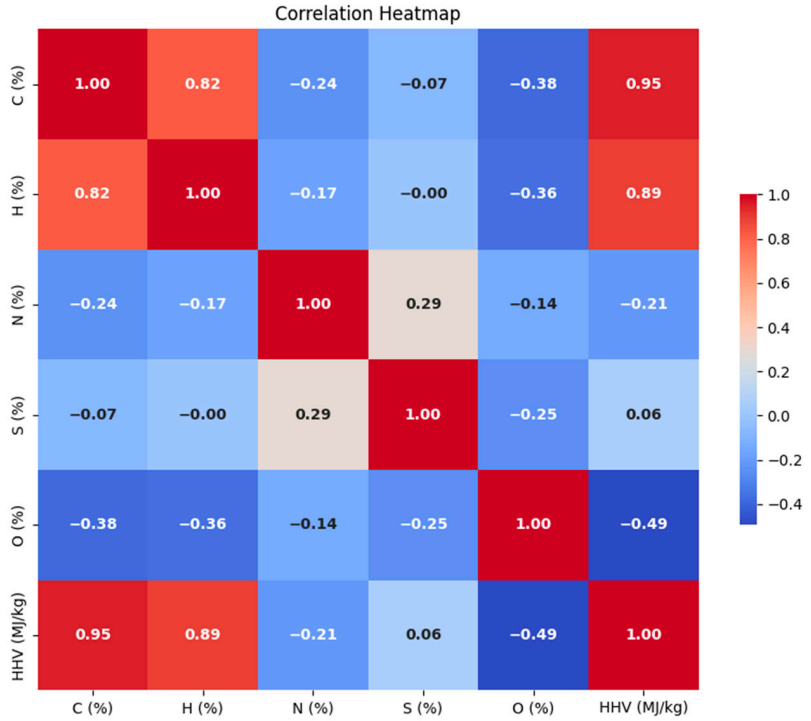

Figure 5. Pearson correlations between input and output variables.

Table 1. Statistical measures of the variables.

Statistics	O (%)	S (%)	N (%)	H (%)	C (%)	Experimental HHV
Min	0.00	0.00	0.00	2.00	9.00	3.50
Max	48.62	2.64	10.00	14.50	92.00	49.30
Mean	31.39	0.31	1.18	6.43	49.09	21.19
Std	13.98	0.47	1.50	2.27	15.57	8.47

5. Model Results

5.1. Cross Validation Results

Machine learning algorithms include numerous hyperparameters. Typically, some of these parameters can be optimized during training, while others, known as superparameters, cannot be optimized this way and must be manually tuned to ensure the model's integrity. In the current investigation, we used five-fold cross-validation (CV) to assess the impact of the SMA on the hyperparameter tuning of the GBRT model. This approach involves randomly dividing the data sets into five groups. Each group is alternately designated as the test group, while the remaining four groups serve as the training groups. The ideal set of hyperparameter combinations was discovered after five cross-validations. Five-fold CV enhances the model's capacity for generalization and prevents both overfitting and underfitting, while optimizing the utilization of the data sets. Table 2 presents the range, default, and ideal settings of the parameters obtained from the hyperparameter tuning methodology. Five parameters were optimized in our tuning process: max_depth, learning_rate, n_estimators, subsample, and min_samples_split. Max_depth limits tree depth to prevent overfitting. Learning_rate adjusts each tree's impact, balancing accuracy and complexity. N_estimators set the number of trees, influencing performance and computation. Subsample uses a fraction of training data for each tree, adding randomness. Min_samples_split ensures nodes have sufficient data before splitting.

Table 2. Range and optimized parameters obtained from the SMA-GBRT.

Parameter	Range	Default	Best Value
max_depth	[3, 10]	3	10.000
learning_rate	[0.01, 0.3]	0.1	0.233
n_estimators	[50, 500]	100	234.000
subsample	[0.1, 1.0]	1	0.675
min_samples_split	[2, 20]	2	2.000

This study chose to analyze the RMSE value, which represents error, and the R^2 value, which represents consistency, separately to further assess the predictive accuracy of the hybrid machine learning model from a quantitative perspective. These two values were computed for both the model training and validation datasets. A lower RMSE indicates better predictive accuracy, while a higher R^2 value signifies a stronger correlation between the predicted and actual outcomes. The RMSE value ranges between 0 and the maximum value of the observed data, while R^2 ranges between 0 and 1. The following formulas were used to calculate the RMSE and R^2 values:

Root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(y_{pi} - y_{Ti}\right)^2}{n}} \tag{21}$$

Correlation of determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_{pi} - y_{Ti}\right)^2}{\sum_{i=1}^{n} \left(\bar{y}_{Ti} - y_{Ti}\right)^2} \tag{22}$$

The formula uses the following notations: n is the sample size, y_{pi} is the predicted value, y_{Ti} is the actual value, $\overline{y}_{Ti}$ is the average of the actual values.

The results of five-fold cross-validation (CV) are displayed in Figure 6. All but the first fold has RMSE values above two, with the third fold having the lowest RMSE value. This indicates that the SMA is capable of successfully adjusting the GBRT hyperparameters, with the third fold yielding the best hyperparameter tuning results. Notably, the prediction performances of the testing set and training set were further examined to ascertain the accuracy of the SMA and GBRT hybrid machine learning model due to the anomalous phenomena of overfitting and underfitting.

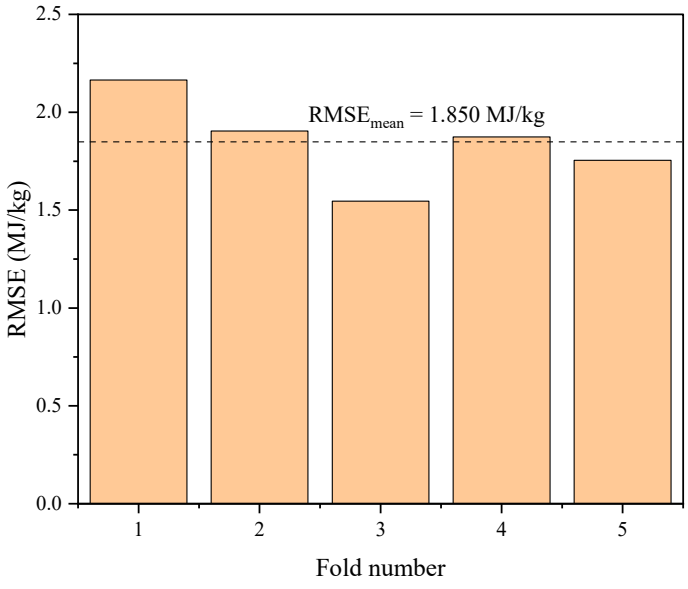

Figure 6. Five-Fold cross-validation RMSE results for the SMA-GBRT model.

5.2. Evaluating the Prediction Performance

The accuracy of the SMA and GBRT hybrid machine learning model in predicting the HHV of MSW was confirmed in this study by analyzing the consistency of the test set and the training set. The hybrid model's prediction performance on the HHV is displayed in Figure 7, where the histogram illustrates the error between the predicted and actual values. It is evident from Figure 7a that there is a nearly perfect match between the predicted and actual numbers. In Figure 7b, there is very little variation between the predicted and actual values for the individual samples, showing a high degree of consistency for the majority of the samples. It is clear that there is sufficient consistency between the measured and anticipated values. The overall consistency between the predicted and measured values is acceptable, indicating that the prediction accuracy of the SMA and GBRT models for the HHV of MSW is also acceptable. Although the testing set has more errors than the training set and some samples show significant deviations between the predicted and actual values, the overall prediction performance remains reliable.

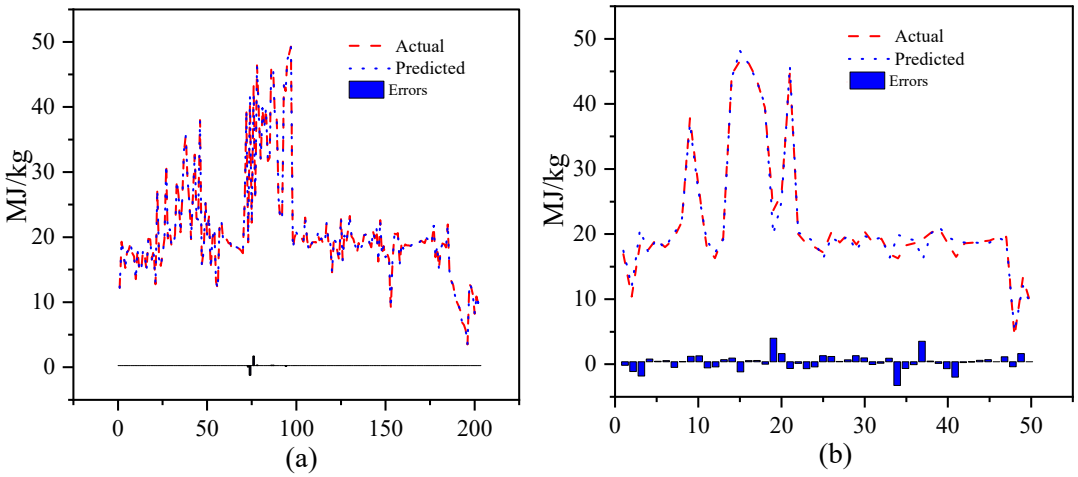

Figure 7. The approximate relationship between the actual and predicted (**a**) training set and (**b**) testing set.

Figure 8 shows that the forecasted and measured values coincide quite well, with the forecasted and observed HHVs mostly concentrated in the range of 4–48 MJ/kg. By quantitatively analyzing the parameters in the training and testing sets—specifically the R^2 values and RMSE values—it is possible to confirm the strong consistency between the predicted and measured values. Both the training set and the test set near perfect R^2 values and very low RMSE values, as evidenced by the data sets' respective R^2 values of 0.999 and 0.984 and their corresponding RMSE values of 0.145 and 1.175 for model training and model effectiveness verification. These findings demonstrate that the proposed hybrid machine learning model of the SMA and GBRT for forecasting the HHV of MSW is free from the overfitting paradox. The reliability of the SMA and GBRT hybrid machine learning model for assessing the HHV of MSW is further confirmed by the good R^2 values and RMSE values.

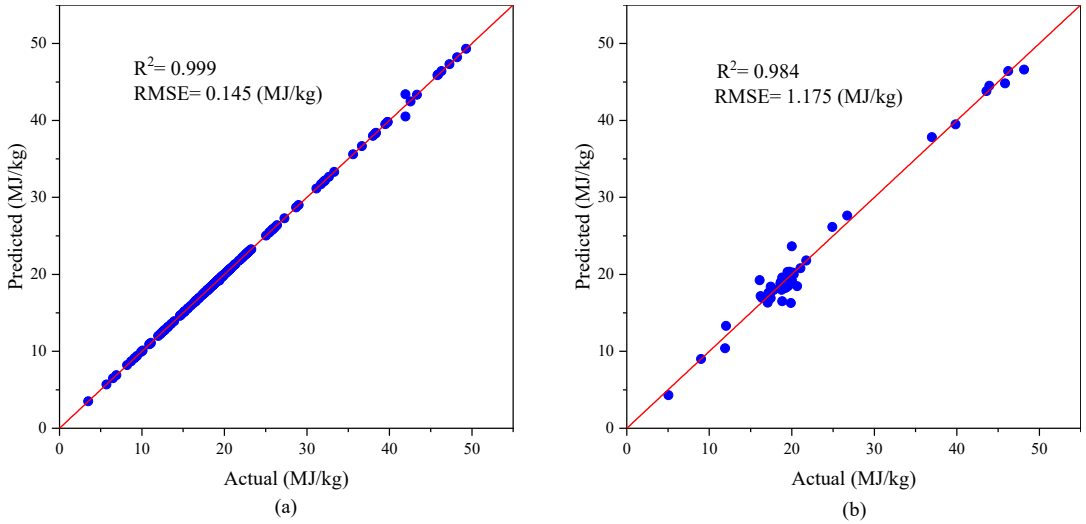

Figure 8. Scatter plots of the actual and predicted (**a**) training set and (**b**) testing set.

5.3. SMA-GBRT and Default GBRT Comparison

The performance comparison between the SMA-GBRT, with parameters optimized by the metaheuristic method, and the GBRT, with default values suggested by the GBRT toolbox, is shown in Table 3. It is evident that the SMA-GBRT outperforms the default GBRT on every statistic during the training phase. Compared to the default GBRT, the RMSE and MAE of the SMA-GBRT are approximately 71.50% and 95.61% lower, respectively. This pattern remains consistent during the testing phase as well. In every statistic, the SMA-GBRT performs better than the default GBRT. Specifically, the RMSE and MAE of the SMA-GBRT are 19.61% and 21.33% lower, respectively, compared to the GBRT. Additionally, during the testing phase, the SMA-GBRT's R^2 significantly outperforms that of the GBRT. This finding suggests that utilizing the SMA to fine-tune parameters helps the SMA-GBRT to reduce overfitting and produce more accurate predictions.

Table 3. The performance comparison between the SMA-GBRT and default GBRT.

Phase	Metrics	Default GBRT	SMA GBRT
Training	RMSE	0.510	0.145
	MAE	0.383	0.017
	R^2	0.996	1.000
Testing	RMSE	1.462	1.175
	MAE	1.036	0.815
	R^2	0.976	0.984

5.4. Reliability Analysis

Our hybrid model GBRT-SMA was compared with other previously developed machine learning models from previous research [24]. The previous research used three machine learning models: Gene Expression Programming (GEP), a Feed-Forward Neural Network (FFNN), and a hybrid Rank-Based Ant System and Support Vector Machine (RBAS-SVM). Gene Expression Programming (GEP) uses populations of individuals selected based on fitness and introduces genetic variation through genetic operators, similar to genetic programming (GP) and genetic algorithms (GAs) [39]. The GEP evolutionary algorithm integrates features from both GP and GA. It utilizes expression parse trees of various shapes and sizes from GP and linear chromosomes of fixed lengths from GA [40]. The GEP formulation produced by previous research is as follows:

$$HHV\left[\frac{MJ}{kg}\right] = -3.772 + 0.0035C^2 + 6.306\sqrt{H} \quad (23)$$

On the other hand, the Feed-Forward Neural Network (FFNN) is a computer system based on the biological neural networks found in animal brains. Although neural networks models are nonlinear regressions, their great versatility stems from their large number of parameters, which enable them to approximate any smooth function. Finally, Support Vector Regression (SVR) is a popular branch of machine learning and has been extensively used for modeling complex nonlinear systems. However, the parameters (C, ε, and kernel parameter γ) significantly impact the SVR prediction performance. If these parameters are not properly tuned, the resulting SVR model may suffer from overfitting or underfitting. To avoid the trial-and-error process in selecting the best SVR input variables, the previous research [24] developed a hybrid model combining the Rank-Based Ant System (RBAS) optimization approach with SVR, known as an RBAS-SVM. We would like to point out that the same training/testing data from the research [24] were used in our comparison of the models' performances. This will prevent any inconsistencies and ensure a fair and accurate evaluation of the model's predictive capabilities.

As shown in Figure 9, performance charts based on R^2 and RMSE were created to measure and compare the prediction performance of these models more easily. The SMA-GBRT model's prediction error is clearly smaller than that of any other model. For training, testing, and all datasets, the SMA-GBRT model exhibits the highest R^2 and the lowest RMSE values. One the other hand, the performance of the GEP equation is worse than that of the other models.

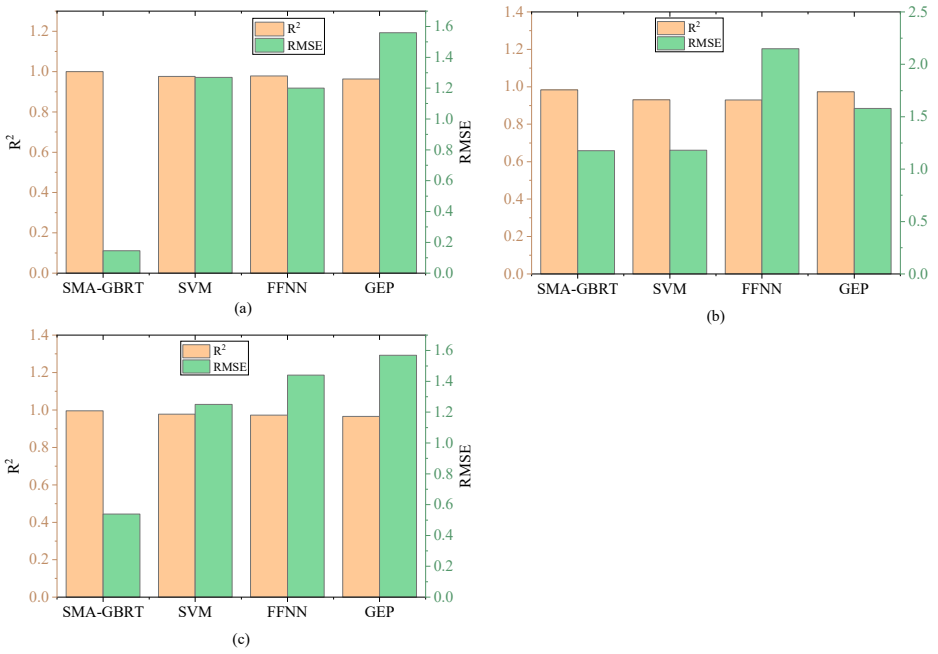

Figure 9. Performance comparison between the model developed in this study and previously developed models for (**a**) the training dataset, (**b**) the testing dataset, and (**c**) the entire dataset.

To understand the accuracy and effectiveness of machine learning models, examining the histogram of residual distributions is crucial. These histograms reveal the discrepancies between predicted and actual values, offering valuable insights. In this context, Figure 10 presents the residual distribution histograms for four models. The residuals of all the models show a pattern consistent with a normal distribution, indicating strong predictive capabilities. Notably, the SMA-GBRT model stands out with consistently lower mean residuals and standard deviations in both training and testing sets, signifying higher predictive stability and accuracy. Specifically, Figure 8c shows that the SMA-GBRT model has an overall mean residual of −0.0147 MJ/kg and a standard deviation of 0.5404. Compared to the SVM, FNN, and GEP models, the SMA-GBRT model reduces standard deviations by 57.02%, 62.18%, and 68.32%, respectively. Thus, the residual distribution histograms clearly

demonstrate that the SMA-GBRT model is the most effective predictive model for the HHV of MSW.

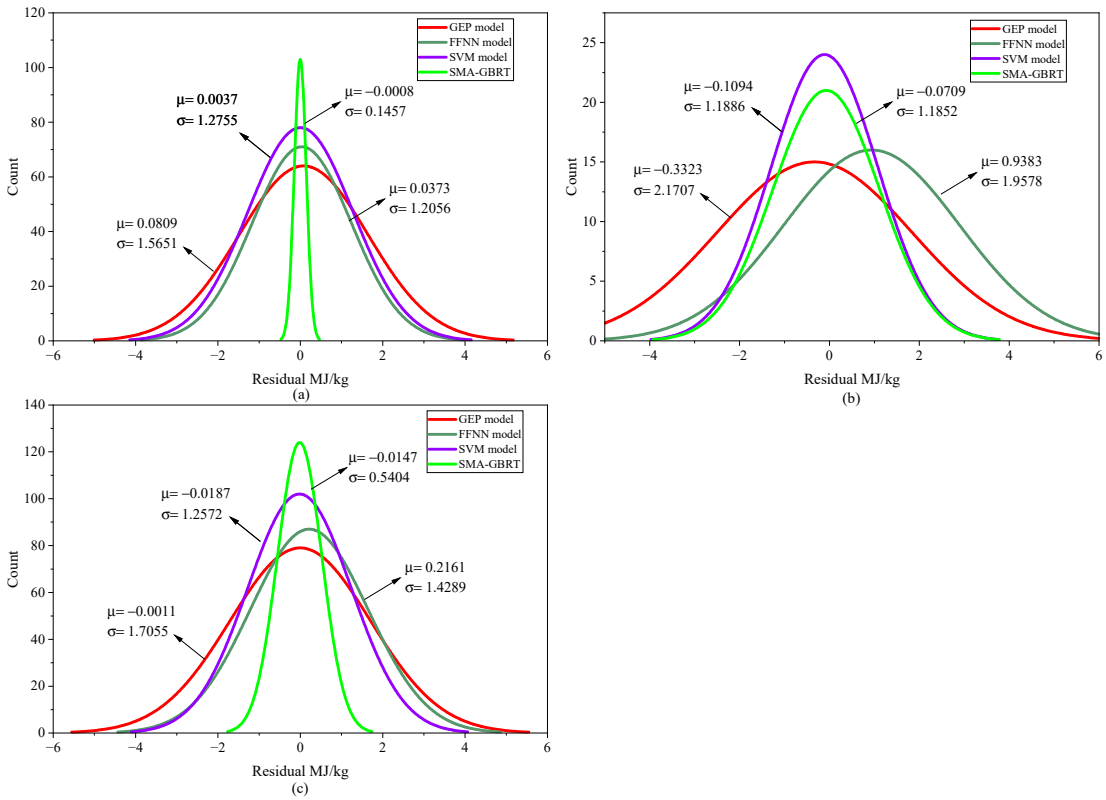

Figure 10. Residual normal distribution for our model and previously developed model: (**a**) training; (**b**) testing; and (**c**) all data.

In general, the prediction accuracy and generalization capacity of single machine learning models (SVM, FFNN, and GEP) from earlier research [24] are limited. A single learner is also vulnerable to overfitting. In this paper, the GBRT, a type of ensemble machine learning with better performance, is proposed as a solution to these problems. To complete classification or regression tasks, it trains and integrates multiple weak learners. Since the ensemble's final product reflects the combined contributions of all the weak learners, its performance is enhanced. Furthermore, the ensemble algorithm's search space is expanded by combining several weaker learners, making it easier to escape local minima. The ensemble model has demonstrated better accuracy and generalization capacity than single machine learning models in numerous experiments. Consequently, it has been extensively applied and has recently produced encouraging outcomes in a range of regression and classification problems.

5.5. Graphical User Interface

Despite the fact that the SMA-GBRT model can produce reliable HHV forecasts, its status as a 'black box' makes it inconvenient for academics and engineers to use in real-world applications. To address this, a web application (WA) has been developed to provide easy access to the SMA-GBRT model for quick HHV prediction. This WA allows users to interactively input numerical values for the C (%), O (%), H (%), N (%), and S (%) parameters. The HHV prediction then appears directly on the screen. Access the SMA-GBRT model web application at https://cujzdggbw59mq5rhropmph.streamlit.app/.

6. Conclusions

Using a wide range of HHVs of MSW-based materials utilizing their contents of carbon, hydrogen, oxygen, nitrogen, and sulfur extracted from the research by [24], Gradient Boosting Regression Tree types of machine learning optimized by the Slime Mould Algorithm were used in this study to decrease the difficulties of experimental calculations, such as the expensive and time-consuming procedures required to predict the HHV of MSW. The following bullet points are the main conclusions from this study:

- The results of the five-fold cross-validation showed that the SMA-GBRT model successfully optimized the RMSE performance metric, achieving values below two for most folds. This indicates a strong generalization capability of the SMA-GBRT model, effectively minimizing the risks of overfitting and underfitting.
- The performance analysis graphs for the training and testing sets showed nearly perfect matches between the predicted and actual HHVs. Additionally, the RMSE and R^2 values, as performance indicators, were exceptionally low and high, respectively, confirming the model's reliability and precision.
- When compared to the default GBRT model and other previously developed machine learning models (GEP, FFNN, and RBAS-SVM), the SMA-GBRT model outperformed them in all statistical metrics. The SMA-GBRT model exhibited lower RMSE and MAE values, as well as higher R^2 values, both in the training and testing phases. This indicates that the SMA optimization significantly enhances the GBRT model's predictive capabilities.
- The residual distribution histograms further substantiated the SMA-GBRT model's superiority, showing lower mean residuals and standard deviations compared to other models. This highlights the model's predictive stability and accuracy.
- To address the practical usability issues of the SMA-GBRT model as a "black box", a web application was developed. This application allows users to input relevant parameters interactively and obtain immediate HHV predictions, thereby making the model accessible and convenient for real-world applications.

7. Futures and Perspectives

Gradient Boosting Regression Trees optimized with the Slime Mould Algorithm have the potential to significantly impact the prediction of the higher heating values of municipal solid waste. This could lead to increased waste management efficiency, improved energy efficiency, and technological innovations.

Furthermore, an accessible interface for the proposed model will contribute to the widespread adoption of the technique and facilitate its implementation in the municipal waste processing industry.

Future plans include expanding the experimental database and utilizing other algorithms to achieve better and more accurate predictions.

Author Contributions: Methodology, E.Q.S.; data curation, F.F.T.; writing—original draft, H.I.; writing—review and editing, S.H.M., K.A.O. and M.P.; project administration, K.A.O.; funding acquisition, K.A.O.; formal analysis, E.Q.S.; software, F.F.T.; conceptualization, S.H.M.; supervision, H.I.; investigation, K.A.O.; and resources, M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in [24].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Azam, M.; Jahromy, S.S.; Raza, W.; Raza, N.; Lee, S.S.; Kim, K.-H.; Winter, F. Status, characterization, and potential utilization of municipal solid waste as renewable energy source: Lahore case study in Pakistan. *Environ. Int.* **2020**, *134*, 105291. [CrossRef]
2. Klemeš, J.J.; Van Fan, Y.; Tan, R.R.; Jiang, P. Minimising the present and future plastic waste, energy and environmental footprints related to COVID-19. *Renew. Sustain. Energy Rev.* **2020**, *127*, 109883. [CrossRef]
3. Dalmo, F.C.; Simão, N.M.; de Lima, H.Q.; Jimenez, A.C.M.; Nebra, S.; Martins, G.; Palacios-Bereche, R.; de Mello Sant'Ana, P.H. Energy recovery overview of municipal solid waste in São Paulo State, Brazil. *J. Clean. Prod.* **2019**, *212*, 461–474. [CrossRef]
4. Lin, B.; Raza, M.Y. Analysis of energy related CO_2 emissions in Pakistan. *J. Clean. Prod.* **2019**, *219*, 981–993. [CrossRef]
5. Tang, S.; Mei, W.; Yang, J.; Wang, Z.; Yang, G. Effect of anaerobic digestion pretreatment on pyrolysis of distillers' grain: Product distribution, kinetics and thermodynamics analysis. *Renew. Energy* **2024**, *221*, 119721.
6. Akpodiogaga-a, P.; Odjugo, O. General overview of climate change impacts in Nigeria. *J. Hum. Ecol.* **2010**, *29*, 47–55. [CrossRef]
7. Azadi, S.; Karimi-Jashni, A. Verifying the performance of artificial neural network and multiple linear regression in predicting the mean seasonal municipal solid waste generation rate: A case study of Fars province, Iran. *Waste Manag.* **2016**, *48*, 14–23. [CrossRef]
8. Kaza, S.; Yao, L.; Bhada-Tata, P.; Van Woerden, F. *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*; World Bank Publications: Chicago, IL, USA, 2018.
9. Iqbal, A.; Zan, F.; Liu, X.; Chen, G.-H. Integrated municipal solid waste management scheme of Hong Kong: A comprehensive analysis in terms of global warming potential and energy use. *J. Clean. Prod.* **2019**, *225*, 1079–1088. [CrossRef]
10. Directive (EU) 2018/850 of the European Parliament and of the Council of 30 May 2018 amending Directive 1999/31/EC on the landfill of waste. *Off. J. Eur. Union* **2018**, *150*, 100–108.
11. Vlaskin, M.S. Municipal solid waste as an alternative energy source. *Proc. Inst. Mech. Eng. Part A J. Power Energy* **2018**, *232*, 961–970. [CrossRef]
12. Guberman, R. What is Waste-to-Energy? RTS. Available online: https://www.rts.com/blog/what-is-waste-to-energy/ (accessed on 19 July 2021).
13. Hockenos, P. EU Climate Ambitions Spell Trouble for Electricity from Burning Waste. Clean Energy Wire. Available online: https://www.cleanenergywire.org/news/eu-climate-ambitions-spell-trouble-electricity-burning-waste (accessed on 12 August 2024).
14. Hockenos, P. Waste to Energy—Controversial Power Generation by Incineration. Available online: https://www.cleanenergywire.org/factsheets/waste-energy-controversial-power-generation-incineration (accessed on 20 July 2024).
15. Pfadt-Trilling, A.R.; Volk, T.A.; Fortier, M.O.P. Climate Change Impacts of Electricity Generated at a Waste-to-Energy Facility. *Environ. Sci. Technol.* **2021**, *55*, 1436–1445. [CrossRef] [PubMed]
16. Trends in Solid Waste Management. World Bank. Available online: https://datatopics.worldbank.org/what-a-waste/trends_in_solid_waste_management.html (accessed on 20 July 2024).
17. Erdoğan, S. LHV and HHV prediction model using regression analysis with the help of bond energies for biodiesel. *Fuel* **2021**, *301*, 121065. [CrossRef]
18. Xing, J.; Luo, K.; Wang, H.; Gao, Z.; Fan, J. A comprehensive study on estimating higher heating value of biomass from proximate and ultimate analysis with machine learning approaches. *Energy* **2019**, *188*, 116077. [CrossRef]
19. Ullah, Z.; Khan, M.; Naqvi, S.R.; Khan, M.N.A.; Farooq, W.; Anjum, M.W.; Yaqub, M.W.; AlMohamadi, H.; Almomani, F. An integrated framework of data-driven, metaheuristic, and mechanistic modeling approach for biomass pyrolysis. *Process Saf. Environ. Prot.* **2022**, *162*, 337–345. [CrossRef]
20. Naveed, M.H.; Khan, M.N.A.; Mukarram, M.; Naqvi, S.R.; Abdullah, A.; Haq, Z.U.; Ullah, H.; Al Mohamadi, H. Cellulosic biomass fermentation for biofuel production: Review of artificial intelligence approaches. *Renew. Sustain. Energy Rev.* **2024**, *189*, 113906. [CrossRef]
21. Taki, M.; Rohani, A. Machine learning models for prediction the Higher Heating Value (HHV) of Municipal Solid Waste (MSW) for waste-to-energy evaluation. *Case Stud. Therm. Eng.* **2022**, *31*, 101823. [CrossRef]
22. Zhu, X.; Yang, G. Study on HHV prediction of municipal solid wastes: A machine learning approach. *Int. J. Energy Res.* **2022**, *46*, 3663–3673. [CrossRef]
23. Ibikunle, R.A.; Lukman, A.F.; Titiladunayo, I.F.; Akeju, E.A.; Dahunsi, S.O. Modeling and robust prediction of high heating values of municipal solid waste based on ultimate analysis. *Energy Sources Part A Recovery Util. Environ. Eff.* **2020**, 1–18. [CrossRef]
24. Bagheri, M.; Esfilar, R.; Golchi, M.S.; Kennedy, C.A. A comparative data mining approach for the prediction of energy recovery potential from various municipal solid waste. *Renew. Sustain. Energy Rev.* **2019**, *116*, 109423. [CrossRef]
25. Olatunji, O.; Akinlabi, S.; Madushele, N.; Adedeji, P.A. Estimation of Municipal Solid Waste (MSW) combustion enthalpy for energy recovery. *EAI Endorsed Trans. Energy Web* **2019**, *6*, e9. [CrossRef]
26. Aladejare, A.E.; Onifade, M.; Lawal, A.I. Application of metaheuristic based artificial neural network and multilinear regression for the prediction of higher heating values of fuels. *Int. J. Coal Prep. Util.* **2022**, *42*, 1830–1851. [CrossRef]
27. Bui, H.-B.; Nguyen, H.; Choi, Y.; Bui, X.-N.; Nguyen-Thoi, T.; Zandi, Y. A novel artificial intelligence technique to estimate the gross calorific value of coal based on meta-heuristic and support vector regression algorithms. *Appl. Sci.* **2019**, *9*, 4868. [CrossRef]
28. Adeleke, O.; Akinlabi, S.; Jen, T.-C.; Adedeji, P.A.; Dunmade, I. Evolutionary-based neuro-fuzzy modelling of combustion enthalpy of municipal solid waste. *Neural Comput. Appl.* **2022**, *34*, 7419–7436. [CrossRef]

29. Oh, E. Prediction of Gasification Process via Random Forest Regression Model Optimized with Meta-heuristic Algorithms. *J. Artif. Intell. Syst. Model.* **2024**, *1*, 45–65.
30. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
31. Shatnawi, A.; Alkassar, H.M.; Al-Abdaly, N.M.; Al-Hamdany, E.A.; Bernardo, L.F.A.; Imran, H. Shear strength prediction of slender steel fiber reinforced concrete beams using a gradient boosting regression tree method. *Buildings* **2022**, *12*, 550. [CrossRef]
32. Sage, A.J.; Genschel, U.; Nettleton, D. Tree aggregation for random forest class probability estimation. *Stat. Anal. Data Min. ASA Data Sci. J.* **2020**, *13*, 134–150. [CrossRef]
33. Blanquero, R.; Carrizosa, E.; Molero-Río, C.; Morales, D.R. Sparsity in optimal randomized classification trees. *Eur. J. Oper. Res.* **2020**, *284*, 255–272. [CrossRef]
34. He, Q.; Kamarianakis, Y.; Jintanakul, K.; Wynter, L. Incident duration prediction with hybrid tree-based quantile regression. In *Advances in Dynamic Network Modeling in Complex Transportation Systems*; Springer: New York, NY, USA, 2013; pp. 287–305.
35. Li, S.; Chen, H.; Wang, M.; Heidari, A.A.; Mirjalili, S. Slime mould algorithm: A new method for stochastic optimization. *Future Gener. Comput. Syst.* **2020**, *111*, 300–323. [CrossRef]
36. Kumar, C.; Raj, T.D.; Premkumar, M.; Raj, T.D. A new stochastic slime mould optimization algorithm for the estimation of solar photovoltaic cell parameters. *Optik* **2020**, *223*, 165277. [CrossRef]
37. Yin, C.-Y. Prediction of higher heating values of biomass from proximate and ultimate analyses. *Fuel* **2011**, *90*, 1128–1132. [CrossRef]
38. Qian, C.; Li, Q.; Zhang, Z.; Wang, X.; Hu, J.; Cao, W. Prediction of higher heating values of biochar from proximate and ultimate analysis. *Fuel* **2020**, *265*, 116925. [CrossRef]
39. Li, P.; Khan, M.A.; Galal, A.M.; Awan, H.H.; Zafar, A.; Javed, M.F.; Khan, M.I.; Qayyum, S.; Malik, M.; Wang, F. Sustainable use of chemically modified tyre rubber in concrete: Machine learning based novel predictive model. *Chem. Phys. Lett.* **2022**, *793*, 139478. [CrossRef]
40. Awoyera, P.O.; Kirgiz, M.S.; Viloria, A.; Ovallos-Gazabon, D. Estimating strength properties of geopolymer self-compacting concrete using machine learning techniques. *J. Mater. Res. Technol.* **2020**, *9*, 9016–9028. [CrossRef]