Applying Support Vector Regression-Based Hybrid Models For Modeling The Gasification Process

Baofu Gong

Library, Criminal Investigation Police University of China, Shen Yang, 110854, China

Corresponding author. E-mail: gongbaofu525@163.com

Received: March 28, 2024; Accepted: June 30, 2024

Modeling the gasification process through machine learning (ML) involves predicting the behavior and performance of gasification systems. Support Vector Regression (SVR) is known as an effective procedure in forecasting continuous variables and is even suitable for the modeling gasification process. Employing optimization algorithms to connect and fine-tune internal model settings can lead to the creation of various hybrid and ensemble models. Generally, employing hybrid models has demonstrated enhanced performance while utilizing cost-effective modeling techniques. In this study, SVR was utilized as a machine learning method, alongside the Crystal Structure Algorithm (CryStAl) and the Population-based Vortex Search Algorithm (PVSA), to fine-tune SVR for accurately assessing CO and CO_2 values. After evaluating the outcomes of the proposed models, it was observed that the SVR-PVSA hybrid model outperformed the SVR-CryStAl model, with differences of 1%, 19%, and 57% based on R^2, RMSE, and MAE indices, respectively for CO and that of 1%, 14%, and 54% for CO_2 in terms of R^2, RMSE, and MAE evaluators, respectively. Furthermore, for predicting both CO and CO_2, the SVR-CryStAl hybrid model yielded the highest value, demonstrating superior performance compared to the SVR-PVSA model, with an average difference of 0.6% and 0.9% in terms of the VAF index.

Keywords: Gasification modeling; Machine learning; Support Vector Regression; Crystal Structure Algorithm; Population-based Vortex Search Algorithm

http://dx.doi.org/10.6180/jase.202506_28(6).0003

1. Introduction

Renewable energy pertains to energy sources that are naturally replenished and do not exhaust the Earth's resources. These sources encompass solar energy, wind energy, hydropower, geothermal energy, and biomass [1]. Various techniques are employed to harness energy from each of these sources. One of the most recent methods used today is a thermochemical process known as gasification [2]. Gasification involves converting carbonaceous materials such as coal, biomass, or waste into synthetic gas (syngas) by subjecting them to high temperatures with a controlled amount of oxygen or steam. These syngas can then be utilized as fuel or as a raw material for producing chemicals and other valuable products. Gasification presents several advantages over traditional combustion processes, as it is more efficient in terms of fuel usage and emits fewer pollutants, thereby offering a cleaner energy alternative [3–5]. While there has been substantial research on the gasification of diverse biomass and waste feedstocks, a comprehensive understanding of the broader environmental, economic, and social implications of this technology is necessary to facilitate its scalable and efficient adoption. To achieve this, precise process models are indispensable. Historically, thermodynamic equilibrium, kinetic, and computational fluid dynamics (CFD) models have been extensively and effectively employed for this purpose [6].

The application of machine learning (ML) holds great

potential in predicting gasification processes. ML has the capability to forecast gasification outputs with greater precision compared to traditional models [6, 7]. Machine learning, a subset of artificial intelligence involves training computers to learn from data and make predictions or take actions without explicit programming. It employs algorithms and statistical models to enable computers to analyze and interpret data patterns and utilize them to perform tasks or make decisions. ML is widely used in various domains, such as image recognition, natural language processing, recommendation systems, and autonomous vehicles [8]. In the context of gasification processes, ML can enhance efficiency, optimize operating conditions, and improve overall performance [9, 10]. By leveraging historical data on gasification processes and influencing factors, ML algorithms can be trained to make accurate predictions. For instance, George et al. developed an Artificial Neural Network (ANN) using a dataset from a lab-scale bubbling fluidized bed gasifier, demonstrating a high-accuracy forecast of syngas production from regional biomass [11]. Similarly, Ozbas et al. (2019) and Shenbagaraj et al. (2021) achieved satisfactory levels of accuracy using ML models in simulating biomass gasification and predicting the composition and yield of syngas, respectively [12, 13]. Li et al. (2021) employed a gradient-boosting model specifically designed for the hydrothermal gasification of waste feedstock types [14].

A recent departure from traditional models involves the utilization of ML algorithms, including Support Vector Regressions (SVRs). These algorithms excel at mapping complex, non-linear relationships between input and output variables [15]. Consequently, they hold the promise of offering more precise predictions of gasification outputs when compared to conventional models. Most studies have demonstrated that SVRs possess the ability to forecast diverse output parameters in various modeling analyses with a notable degree of prediction accuracy. Ullah et. al., in a study used a SVR for prediction of adsorption capacity of Fe-modified biochar for selenium that appeared highly effective model compared to random-forest approach like another study conducted by Zaghloul et. al., to model aerobic granular sludge reactors that results showed successful performance of SVR [16, 17]. SVR is a variant of the Support Vector Machines (SVM) algorithm that is used for regression tasks [15]. In this investigation, the SVR model was employed for gasification prediction. By training SVR on historical data related to gasification, it can learn patterns and relationships in the data and predict the gasification output for new, unseen data [18, 19].

Prior research has pointed out the limited applicability of existing ML-based gasification models, which struggle to accommodate various feedstocks, gasifying agents, and reactor configurations [6]. Consequently, the objective of this study is to enhance the current state of knowledge by training a robust feedforward SVR model using a comprehensive dataset encompassing a diverse range of feedstocks and operational scenarios. The innovation of the work is to develop a hybrid model in which two optimizers, the Population-based Vortex Search Algorithm [20] and Crystal Structure Algorithm [21], are employed to be combined with SVR.

In mechanical and gasification processes, single models are mainly used, and hybrid models are rarely found in the literature. The developed model is exceptionally versatile and, to the best understanding suitable for a broad spectrum of feedstock varieties (including plastics, sewage sludge, woody and herbaceous biomass, and municipal solid waste); gasifying participants (air, oxygen, and steam, and their combinations); and reactor choices (fluidized bed and fixed bed). Its adaptability is optimized through the innovative incorporation of categorical data, ensuring its practicality in general gasification process design. In fact, for the present article, ML techniques in gasification prediction encompass process optimization, fault detection and diagnosis, control strategy optimization, and real-time monitoring and prediction [22]. The various steps involved in prediction using ML include data collection, data preprocessing, feature selection, model selection, model training, model evaluation, model deployment, and model monitoring and refinement [23].

Categorical variables, such as gasifying agent, reactor type, and bed material, enable the inclusion of crucial gasification factors often overlooked by previous models. To further enhance its utility in general process design, the model has been designed to predict a wide range of output parameters essential for subsequent modeling tasks, including life cycle sustainability assessment (LCSA), which evaluates a system's environmental, economic, and social impacts. This, in turn, empowers policymakers and investors with the knowledge of the broader system-wide effects, enabling them to make informed decisions.

2. Materials and methods

Initially, the data set for gasification was gathered, consisting of 312 samples encompassing factors such as feedstock composition, temperature, steam/biomass, gas flow rate, and other pertinent parameters obtained from the literature [24–52] to train the proposed SVR model (Table 1). Care was taken to ensure that the data set had minimal missing or erroneous values. Actually, in case of missing

Table 1. The statistic properties of the input variable of CO-CO_2.

Variables	Indicators				
	Category	Min	Max	Avg.	St. Dev
Particle size [mm]	Input	0.250	70.000	5.570	6.681
LHV [MJ/kg wb]	Input	11.500	42.900	18.437	6.549
Carbon [%daf]	Input	40.070	86.034	51.423	9.232
Hydrogen [%daf]	Input	3.792	14.228	6.864	1.997
Nitrogen [%daf]	Input	0.040	7.321	1.116	1.850
Sulphur [%daf]	Input	0.010	1.607	0.427	0.442
Oxygen [%daf]	Input	4.621	53.400	41.337	6.867
Ash [%db]	Input	0.270	44.000	7.433	11.898
Moisture [%wxb]	Input	0.200	27.000	9.138	4.999
Volatile Matter [%db]	Input	56.000	89.110	77.404	8.279
Fixed carbon [%db]	Input	9.067	23.820	15.258	2.775
Cellulose [%db]	Input	29.600	46.200	39.254	1.750
Hemicellulose [%db]	Input	14.000	29.600	20.973	1.576
Lignin [%db]	Input	14.000	27.800	23.676	1.257
Temperature [°C]	Input	553.000	1050.0	799.128	80.876
Residence time [min]	Input	6.000	403.0	109.143	80.230
Steam/Biomass [wt/wt]	Input	0.000	4.040	1.181	0.616
ER [-]	Input	0.090	0.870	0.300	0.093
CO[vol.%db]	Output	2.200	50.000	19.430	9.540
CO_2[vol. %db]	Output	0.000	38.250	16.095	6.420

data for an experiment, the average value for the lost data was calculated based on the available parameters for the same experiment. Subsequently, the significant features influencing gasification process prediction were identified. This involved cleaning the data by removing outliers, addressing missing values, and normalizing or scaling the features. The data set was then split into training and testing subsets. According to the routine in modeling, 70 percent of data was considered for the training stage and 30 percent for testing. The training set was utilized to train the SVR model, while the testing set was used to assess its performance. The SVR model was trained using the training dataset, aiming to identify an optimal hyperplane in a higher-dimensional space for making predictions. This was achieved by maximizing the margin around the hyperplane while allowing for some tolerance for data points to fall within the margin or even on the wrong side of the hyperplane. The trained SVR model was validated using the testing dataset, and its performance was evaluated using appropriate metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), or R-squared (R^2) to gauge its accuracy and predictive capability. Subsequently, the best hyperparameters were identified, and the SVR model was retrained using the entire dataset (both training and testing). The model's performance was evaluated on this final set to obtain an accurate representation of its predictive power. Finally, the trained SVR model was utilized to make predictions on new, unseen data. The input variables were prepared in the same manner as before, and the model was employed to generate predictions for the gasification process variables of interest. Fig. 1 displays the input and output variables along with the corresponding plots. Based on the vast range of datasets available in the literature, releasing CO and CO_2 due to the importance of environmental issues were variables of interest for modeling. As well as modeling these gases are rarely considered in models.

2.1. Support Vector Regression (SVR)

The Support Vector Machine (SVM) algorithm is a widely used method for regression and classification, as per statistical learning theory [53]. SVR was developed to address regression problems by introducing Vapnik's ε-insensitive loss function. In the context of a training dataset containing inputs denoted as X and their corresponding gasification performance outputs labeled as Y, the objective of the SVR model is to find a function, denoted as f(X), that accurately predicts the output Y. This function f(X) is defined as:

$$f(X) = w^T * \Phi(X) + b \quad (1)$$

Here, w represents the weight vector, $\Phi(X)$ is a feature mapping function responsible for converting the input data into a higher-dimensional feature space, and b is the bias term. The feature mapping function $\Phi(X)$ can be constructed using different kernel functions, such as linear, polynomial, radial basis function (RBF), or sigmoid kernels. The choice of kernel function depends on the specific characteristics of the gasification data and the desired modeling

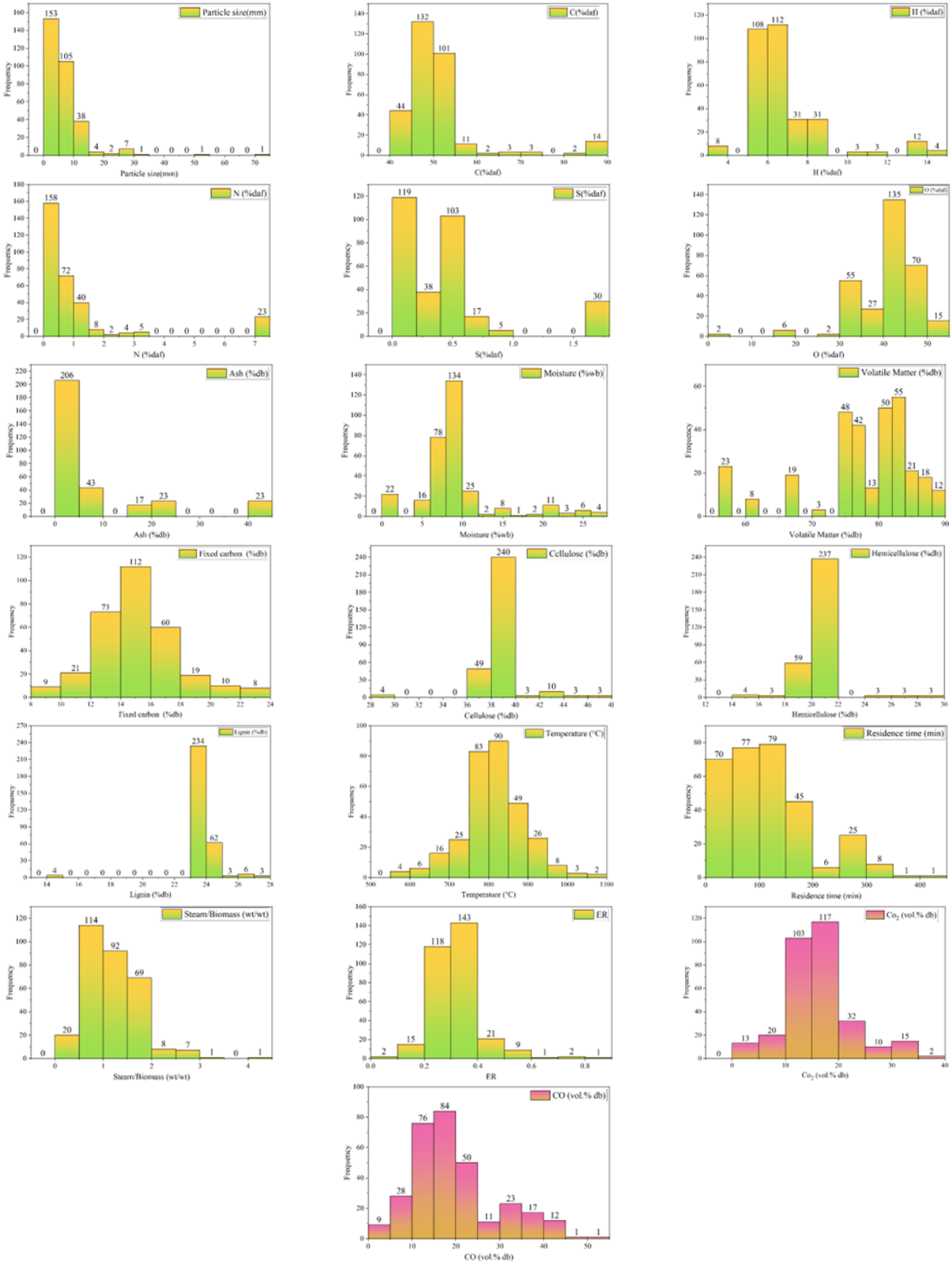

Fig. 1. Input and target data for modeling the gasification process.

accuracy. The SVR model with Gaussian kernel can be introduced as follows:

$$\text{Minimize} \quad \left(\frac{1}{2} \times \|w\|^2 + C \times \sum (\xi_i + \xi_i^*) \right. \tag{2}$$

$$\begin{aligned} &\text{Subject to } Y_i - w^T \times \Phi(X_i) - b \leq \varepsilon + \xi_i \times w^T \\ &* \Phi(X_i) + b - Y_i \leq \varepsilon + \xi_i \xi_i, \xi_i^* \geq 0 \end{aligned} \tag{3}$$

Here, ξ_i and ξ_i^* are slack variables introduced to account for a certain level of error, ε represents the margin of tolerance, and C is a regularization parameter that controls the tradeoff between model complexity and error. The optimization problem's goal is to minimize the error $(\|w\|^2)$ while also penalizing instances that fall outside the margin of tolerance (ξ_i and ξ_i^*). The parameter C plays a role in determining the relative importance of error and margin violations within the objective function. After solving the optimization problem, the SVR model can be effectively utilized to predict gasification performance for new input data by applying it to the function

$$f(X) = w^T \times \Phi(X) + b \tag{4}$$

2.2. Population-based Vortex Search Algorithm (PVSA)

The Population-based Vortex Search Algorithm (PVSA) is a metaheuristic optimization approach inspired by the natural occurrence of vortex formation. This algorithm is designed to tackle intricate optimization problems by emulating the behavior of vortices in fluid dynamics. PVSA maintains a population of candidate solutions, where each solution represents a potential answer to the optimization problem. The algorithm continually updates this population by simulating the movement and interaction of vortices. The motion of vortices in PVSA adheres to the following formula:

$$\begin{aligned} &V_i(t+1) = V_i(t) + A * (X_{\text{best}} - X_i(t)) + B \\ &* (X_{\text{localbest}} - X_i(t)) + C * \left(X_{\text{globalbest}} - X_i(t)\right) \end{aligned} \tag{5}$$

In this equation:

- $V_i(t)$ represents the velocity of the ith vortex at the time t.
- $X_i(t)$ represents the current position of the i^{th} vortex.
- X_{best} represents the best position discovered by any vortex in the population.
- $X_{\text{localbest}}$ represents the best position found by neighboring vortices.
- $X_{\text{globalbest}}$ represents the best position located by any vortex globally.

The coefficients A, B, and C determine how different factors influence the movement of vortices. These coefficients are typically set based on empirical knowledge or through experimentation. The movement equation is employed to update the position of each vortex in the population:

$$X_i(t+1) = X_i(t) + V_i(t+1) \tag{6}$$

Subsequently, the algorithm assesses the fitness of each vortex based on the objective function of the optimization problem. These fitness values are crucial for identifying the best positions found by individual vortices and the globally best position. PVSA continues to iterate and refine the population until a termination condition is met, such as reaching a maximum number of iterations or achieving a satisfactory solution. By simulating the movement and interaction of vortices, the PVSA algorithm efficiently explores the search space, making it capable of discovering near-optimal solutions for complex optimization problems. It is worth noting that the specific implementation details of PVSA, such as population initialization, parameter selection, and termination criteria, may vary based on the problem at hand and user preferences.

2.3. Crystal Structure Algorithm (CryStAl)

The repetition of atoms or molecules in space creates a crystal, which can be made up of a single atom or multiple atoms, depending on its complexity. The term "crystal" originated from the Greek word "krustallos," which refers to "rock crystal" and "ice." The study of crystals is known as crystallography. Common examples of crystals include salt (sodium chloride or halite crystals), sugar (sucrose), and snowflakes. Gemstones like quartz and diamonds are also considered crystals. A crystal has a distinct structure due to its three-dimensional network of repeated units. Large crystals have well-defined flat angles and areas (faces), while crystals without flat faces are known as rock crystals. The lattice is the essential and crucial element of crystals, representing the regular arrangement of points in space. Another key component is the basis, which determines the positions of atoms within the crystals. These 2 components come together to describe the overall arrangement of a crystal. To mathematically represent the crystal's configuration in space, a vector is used to determine the lattice points precisely. This vector assumes an infinite lattice shape, allowing for an accurate depiction of the crystal's arrangement.

$$r = \sum n_i a_i \tag{7}$$

The vector a_i represents the vector that is the shortest while following the main crystallographic directions, while η_i is an integer quantity, and i indicates the total number

of corners made of crystal. These characteristics of crystals serve as the foundational concept for the Crystal Structure Algorithm (CryStAl), which is a metaheuristic algorithm that was just created. The startup procedure of CryStAl's first phase is carried out to determine each solution candidate as a crystal in space:

$$cr = \begin{bmatrix} cr_1 \\ cr_2 \\ \vdots \\ cr_i \\ \vdots \\ cr_n \end{bmatrix} = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^j & \cdots & x_1^d \\ x_2^1 & x_2^2 & \cdots & x_2^j & \cdots & x_2^d \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_i^1 & x_i^2 & \cdots & x_i^j & \cdots & x_i^d \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^1 & x_n^2 & \cdots & x_i^j & \cdots & x_n^d \end{bmatrix}, \quad \begin{cases} i = 1.2....n \\ j = 1.2.....d \end{cases} \tag{8}$$

$$\mathrm{X_i^j} = \mathrm{X_{i,min}^j} + \xi \left(\mathrm{X_{i,max}^j} - \mathrm{X_{i,min}^j} \right) \begin{cases} i = 1.2....n \\ j = 1.2.....d \end{cases} \tag{9}$$

The optimization problem's dimension is represented by d, while the initial number of applicants is indicated by n.$X_j^j(0)$ represents the decision variables' starting values, and $X^j{}_{max}$ are the variables' lower and higher boundaries inside the field of search space. A randomly produced digit in the interval $[0, 1]$ is denoted by ξ.

The main loop of CryStAl is where the process of updating positions for the crystals (candidates) in the search space is carried out. This process utilizes both basic and advanced aspects of crystallography. The corner points in the crystal (Cr_b) are randomly selected to determine the paramount crystals, as well as the identification of the crystal with the optimal arrangement. Additionally, the average of the primary crystals chosen at random is calculated as $\mathrm{F_c}$.

To update the position, the cubic crystal system's common varieties function as a notion for inspiration, and the subsequent formulas are employed:

Simple cubicle:

$$Cr_{\mathrm{new}} + Cr_{\mathrm{old}} + rCr_{\mathrm{main}} \tag{10}$$

Cubicle with the best crystals:

$$Cr_{\mathrm{new}} + Cr_{\mathrm{old}} + r_1 Cr_{\mathrm{main}} + r_2 Cr_b \tag{11}$$

Cubicle with the mean crystals:

$$Cr_{\mathrm{new}} + Cr_{\mathrm{old}} + r_1 Cr_{\mathrm{main}} + r_2 F_c \tag{12}$$

Cubicle with the best and mean crystals:

$$Cr_{\mathrm{new}} + Cr_{\mathrm{old}} + r_1 Cr_{\mathrm{main}} + r_2 Cr_b + r_3 F_c \tag{13}$$

To determine the new position vector for the crystals, $\mathrm{Cr_{new}}$ is used, while $\mathrm{Cr_{rold}}$ represents the previous position vector. $\mathbf{r}_1, \mathrm{r}_2$, and r_3 are three arbitrarily generated digits within $[0, 1]$. The algorithm terminates after reaching a predefined maximum number of iterations or function evaluations. A boundary control flag is also set for variables that are not inside the set boundaries.

2.4. Preparing hybrid models

The construction of hybrid frameworks necessitates the optimization of estimative models concerning internal settings. In this study, SVR incorporates certain arbitrary variables, as model fits are optimized through the utilization of three parameters: ε, ξ, and C. Through the integration of PVSA and CryStAl with SVR, two hybrid models, namely SVPV and SVCS, were formulated. In the modeling stage, the iteration number was considered 200. The parameter of $\mathrm{N_{PoR}}$ (Network Point of Presence) equals 50, the cache size was 200, and the upper and lower bond of the optimization process was determined to be 0.001 and 999 , respectively.

2.5. Performance evaluation methods

The article assesses the predictive accuracy of PVSA-SVR, CryStAl-SVR, and SVR models using various statistical metrics, including the correlation coefficient $\left(\mathrm{R}^2\right)$, mean absolute error (MAE), root mean squared error (RMSE), variance account factor (VAF), and objective detection metrics (OBJ). These statistical evaluators are presented as follows:

$$R^2 = \left(\frac{\sum_{i=1}^{n} (p_i - \bar{p})(r_i - \bar{r})}{\sqrt{\left[\sum_{i=1}^{n} (p_i - p)^2\right]\left[\sum_{i=1}^{n} (r_i - \bar{r})^2\right]}} \right)^2 \tag{14}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (r_i - p_i)^2} \tag{15}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |r_i - p_i| \tag{16}$$

$$VAF = \left(1 - \frac{\mathrm{var}(t_n - e_n)}{\mathrm{var}(t_n)} \right) \times 100 \tag{17}$$

$$OBJ = \left(\frac{t_{\mathrm{train}} - t_{\mathrm{test}}}{et_{\mathrm{train}} + t_{\mathrm{test}}} \right) \frac{MAE_{\mathrm{test}} + RMSE_{\mathrm{train}}}{1 + R^2_{\mathrm{train}}} + \left(\frac{2t_{\mathrm{train}}}{t_{\mathrm{test}} + t_{\mathrm{train}}} \right) \frac{RMSE_{\mathrm{test}} - MAE_{\mathrm{test}}}{1 + R^2_{\mathrm{test}}} \tag{18}$$

In these expressions, p and r represent the corresponding experimental and predicted values for CO and CO_2. The variable n denotes the number of data instances, $\bar{p}$ signifies the average experimental values for CO and CO_2, and $\bar{r}$ denotes the average predicted values for CO and CO_2. To mitigate bias associated with randomly partitioning a dataset into training and testing sets, researchers often employ a k-fold cross-validation procedure. In this method, the dataset is divided into k partitions $(\mathrm{k} = 10)$,

each termed a fold. During each iteration of k-fold cross-validation, one-fold serves as the test set, while the remaining $k-1$ folds are employed for training the model. This process is repeated k times, each with a different test set. Through cross-validation, the standard deviation and mean of key performance indicators are computed.

3. Results and discussion

The SVPV and SVCS hybrid models, along with the SVR predictive main model, were provided with input data for training and validation, resulting in the generation of CO and CO_2 rates as model outputs. The functionality of the models is crucial to investigate, and this section aims to accomplish this task. Accordingly, cross-validation results was introduced using Tables 2 and 3 presents the indicators' results, evaluating the performance of the developed models in modeling the CO and CO_2 values in the sections of training, validation, testing, and all data on CO and CO_2. The evaluators introduced in the previous section compare the models based on the highest and lowest values. Except for R^2 and VAF, the lowest value of the other evaluators indicates proper performance.

Table 2. The cross-validation results of modeling in developed models.

Feature	Model	R^2	RMSE
CO	SVPV	0.974	1.651
	SVCS	0.964	1.846
	SVR	0.957	2.597
CO_2	SVPV	0.968	1.388
	SVCS	0.955	1.597
	SVR	0.948	2.357

The models' performance in the CO section is outlined as follows:

- The hybrid SVPV model achieved the highest R^2 value of 0.983, with only a 2% difference compared to SVR. Additionally, the SVCS hybrid model performed even better, with a difference of only 1% compared to the SVPV model.
- The following assessment, which is connected to RMSE and the most satisfactory outcome, remains SVPV, with a measure of 1.258. In comparison to other models, this value is notable. SVPV displays a 45% variance compared to SVR and a 19% distinction from SVCS.
- MAE is an alternative error descriptor for models. Similar to previous evaluators, in this evaluator, the lowest value is 0.564, which belongs to SVPV. When comparing the model with the other models, SVR and SVCS have values of 1.209 and 1.301, respectively, and their difference from SVPV is 53% and 57%, respectively.
- The VAF, like R^2, must achieve the highest value. The models demonstrate almost identical performances in their respective evaluation sets. However, generally, SVCS exhibits the highest value, at 99.180. Compared to SVR, which performs as low as 95.812, there is a 3% difference. This variance suggests that all models operate effectively in the VAF evaluator.
- For OBJ, the goal is to have the lowest value. In this context, the SVPV hybrid model attains the minimum value of 0.910 in OBJ. In the context of CO_2, the performance of SVR and combined models has been determined as follows:
- The SVPV hybrid model could attain the highest R^2 value of 0.972, leading to disparities of 2% and 1% between SVR and SVCS, respectively.
- In the evaluation of RMSE and optimal performance, SVPV emerged with a value of 1.088, exhibiting a difference of 41% with SVR and 14% with SVCS.
- Another metric to consider is MAE. Similar to the preceding evaluation, in this case, SVPV records the lowest value at 0.479. Contrasting this with other models, SVR and SVCS show values of 0.892 and 1.044, respectively, resulting in variances from SVPV of 46% and 54%, respectively.
- Within the VAF evaluation, the hybrid model SVCS secured the top value, while SVR garnered the lowest among all models, with a difference of around 5%.
- In the context of OBJ, the SVPV model can register a value as minimal as 0.679. This indicates a variance of 57% with SVR and 42% with SVCS, representing the distinctions between these models.

Overall, it can be asserted that SVPV attained the most favorable values for both CO and CO_2, contrasting with SVR, which secured the least favorable values. As a result, the SVPV composite model displayed superior performance, while SVR exhibited the weakest performance. In a study, Baruah et al. (2017) created an artificial neural network (ANN) model for forecasting the yields of H_2, CO, CO_2, and CH_4 from a downdraft gasifier [54]. The model demonstrated outstanding prediction performance, achieving coefficients of determination (R^2) exceeding 0.98 and a mean relative error of 2.65% in contrast to the experimental

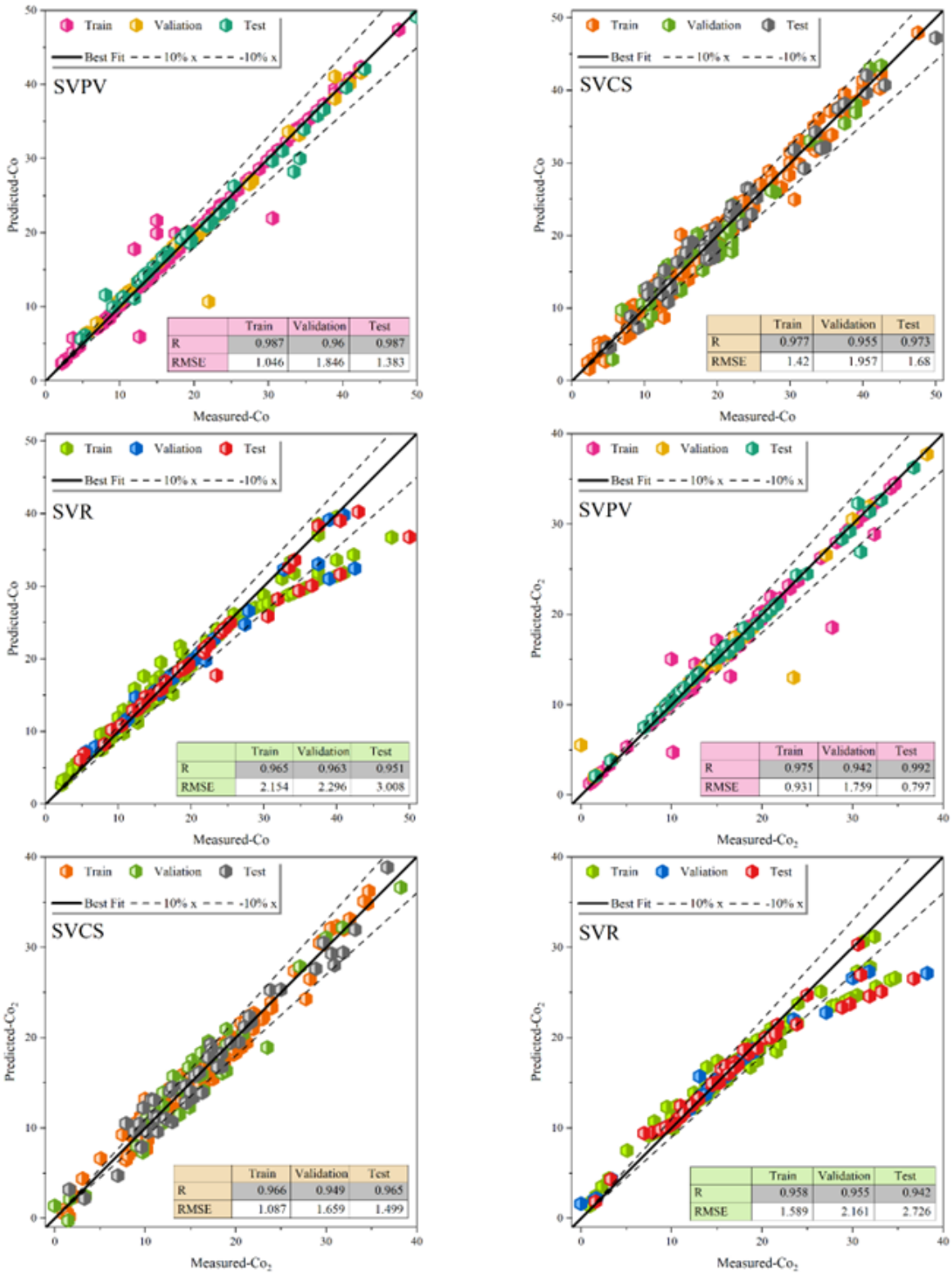

Fig. 2. Plotting the dispersion of evolved hybrid models.

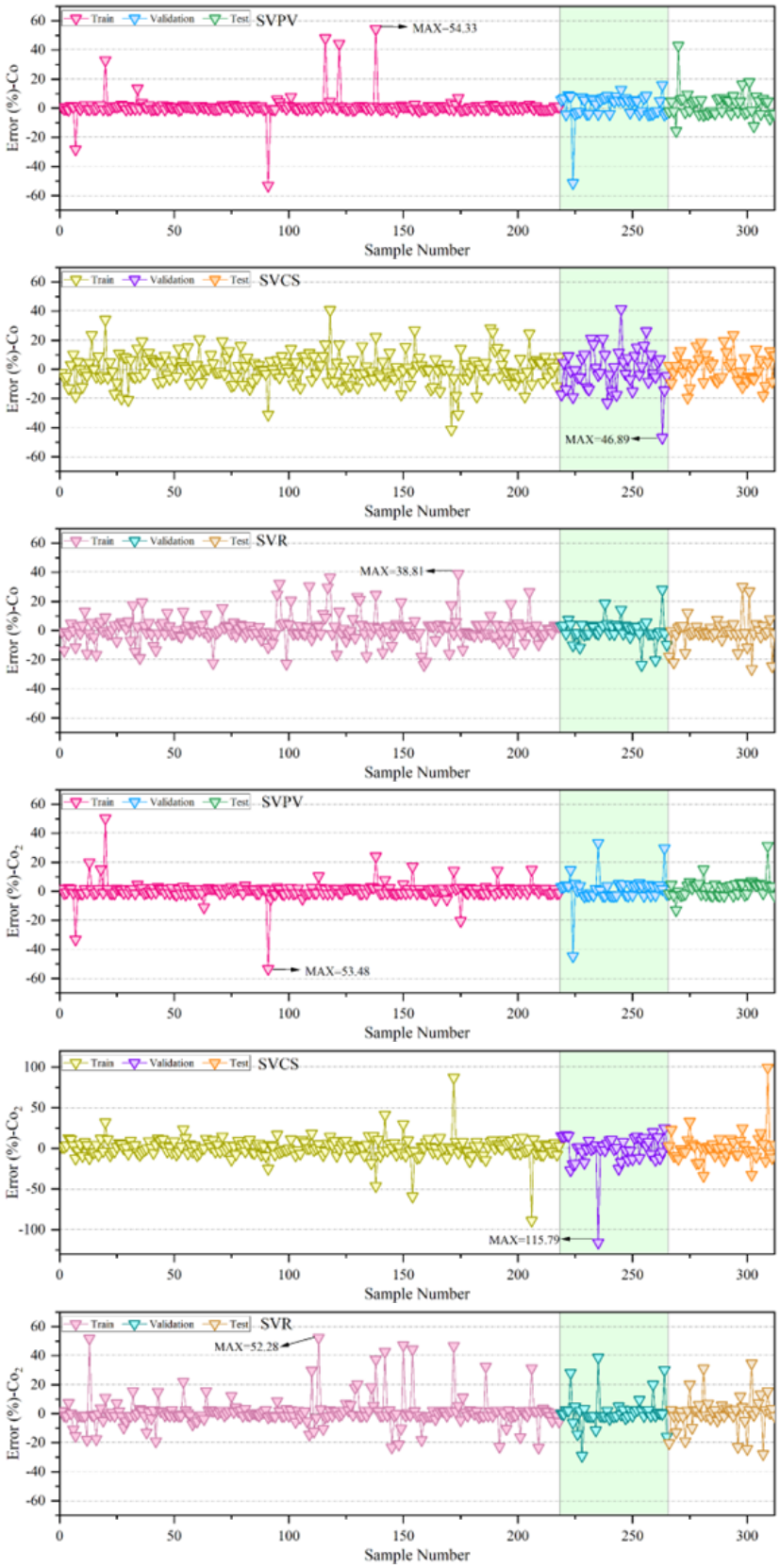

Fig. 3. The error percentage of the models is based on the symbol-line plot.

Table 3. The statistic properties of the input variable of SVR.

	Model name	Modelling phase	Index values				
			R^2	*RMSE*	MAE	*VAF*	*OBJ*
CO	SVPV	Train	0.988	1.047	0.334	98.884	0.910
		Validation	0.960	1.846	1.094	97.394	
		Test	0.987	1.383	1.076	99.291	
		All	0.983	1.258	0.564	98.605	
	SVCS	Train	0.977	1.420	1.164	99.248	1.423
		Validation	0.955	1.957	1.719	98.970	
		Test	0.974	1.681	1.473	99.383	
		All	0.973	1.561	1.301	99.180	
	SVR	Train	0.965	2.154	1.172	96.294	1.952
		Validation	0.964	2.297	1.180	95.426	
		Test	0.952	3.009	1.569	93.807	
		All	0.963	2.294	1.209	95.812	
CO_2	SVPV	Train	0.975	0.931	0.387	97.915	0.679
		Validation	0.943	1.759	0.781	95.273	
		Test	0.992	0.797	0.589	99.555	
		All	0.972	1.088	0.479	97.677	
	SVCS	Train	0.966	1.087	0.919	99.022	1.177
		Validation	0.950	1.660	1.361	98.283	
		Test	0.965	1.500	1.280	99.056	
		All	0.962	1.262	1.044	98.777	
	SVR	Train	0.958	1.589	0.782	94.433	1.589
		Validation	0.956	2.161	1.008	93.045	
		Test	0.942	2.727	1.473	91.865	
		All	0.953	1.854	0.892	93.573	

results. Also, Ascher et al., in 2022, developed a highly generalizable artificial neural network (ANN) model with outstanding prediction accuracy (R^2 = 0.9310 and RMSE = 0.1307) for modeling biomass and waste gasification [6]. Based on modeling results, the correlation rate between observed and simulated values in developed models are in high level that shows a great potentials to modeling. The coefficient of determination in the highest rate was 0.992 that indicates a five percent difference domain comparing to the lowest R^2. By such high-accuracy modeling procedures, the gasification experiments can play a role of physical experiments. However the limitation of using too distant values compared with training data can enter uncertainty to such models.

Fig. 2 illustrates the scatter plot depicting the measured and forecasted levels of CO and CO_2. The scatter plots are determined based on RMSE and R^2. R^2 is responsible for positioning the sample point on the best-fit line. Additionally, the RMSE feature indicates the dispersion of the scatter plots, with lower component values representing lower dispersion and vice versa. As mentioned, the other points are directly correlated with the component and should alternatively have low and high RMSE and R^2 values. A comparison of the two hybrid models, SVPV and SVCS, is shown in Fig. 2a, Fig. 2d, and Fig. 2b, Fig. 2e. The points exhibit different behaviors due to the differing values of the two models. Fig. 2c and Fig. 2f depict the SVR model. The most significant contrast among these three models lies in the RMSE, which demonstrates the dispersion. Fig. 2a and Fig. 2d, associated with SVPV, have lower RMSE values close to 1, indicating much lower dispersion. In general, SVPV displayed better performance in Fig. 2. Also, in the overall view of fitting-line plots in most of models we can see the substantial features to overcome errors in the simulating process. In fact, the performance of Vortex Search Algorithm in tunning the SVR model was more successful than Crystal Structure Algorithm in simulating CO and CO_2 production.

Fig. 3 displays the error percentages of CO and CO_2, evaluated across three phases: training, validation, and testing, using a symbol-line plot. Focusing on SVPV, Fig. 3a and Fig. 3d show that the highest error during the training phase was approximately 54.33% and 53.48% for CO and CO_2, respectively, which decreased to 43 and 44% in the test phase. For SVCS, Fig. 3b and Fig. 3e reveal that the maximum error was equal to 46.89% and 115.79% in the validation phase, respectively. However, the error percentages improved and decreased to 23 and 102% in the testing phase. The improvement in performance during the testing phase suggests that the model has been able to handle the samples better in the training phase, leading to better performance during the testing phase. Finally, Fig. 3c

and Fig. 3f introduce SVR, with the highest error percentage obtained during the training phase being 38.81% and 52.28% for CO and CO_2, respectively, which decreased to 32 and 40% in the test phase. Overall, the SVPV model demonstrated the most suitable and strongest performance compared to other models. Actually, the Vortex Search Algorithm has tried to minimize the bias of modeling we can understand this fact given the plots in Fig. 3 in which most of the experiments have had positions adjacent to the zero line. However, in some experiments vast error rates can be found. On the other hand, for the single SVR model and hybrid model coupled with the Crystal Structure Algorithm, most of the modeled experiments are in the error range of +20% to -20%.

4. Conclusion

This study introduced two hybrid intelligent models and a single model for predicting CO and CO_2 concentrations in producer gas during experimental gasification. Performing multiple tests to measure CO and CO_2 levels represents a significant waste of time and energy. Therefore, the use of artificial intelligence is proposed to estimate CO and CO_2 concentrations through machine learning techniques. The article employs various methods, including SVR as a machine learning approach, along with the Crystal Structure Algorithm (CryStAl) and the Population-based Vortex Search Algorithm (PVSA) as optimizers. SVR was specifically designed to address regression issues by introducing Vapnik's ε-insensitive loss function. The models are examined in two modes—single and hybrid—where hybrid models consist of SVR + CryStAl and SVR + PVSA. In terms of CO, the SVPV hybrid model achieved the highest R^2 value at 0.983, indicating a 2% improvement by SVR. Additionally, compared to the hybrid model, SVCS showed a 1% improvement. When considering RMSE, the SVPV model performed the best with a value of 1.258, marking a 45% difference from SVR and a 19% difference from SVCS. Regarding CO_2, the SVPV hybrid model obtained an R^2 value of 0.972, with a 2% and 1% difference from SVR and SVCS, respectively. The SVPV model also demonstrated the most acceptable performance with an RMSE value of 1.088, showing a 41% difference from SVR and a 14% difference from SVCS. In summary, for both CO and CO_2, the SVPV model yielded the most favorable results, while SVR showed the least satisfactory performance. Consequently, the SVPV composite model exhibited superior performance, while SVR displayed weaker performance. The newly developed model has the capability to forecast gasification outputs for several alternatives for the reactor, gasifying agent, and feedstock. As a result, this eventually lessens the need for expensive and time-consuming experimental investigations since investors and policymakers can quickly and easily examine a wide range of system choices and their performance. Furthermore, the developed model can be integrated with techniques like life cycle sustainability assessment, cost-benefit analysis, and multi-objective optimization. This allows the model to predict many important gasification outputs, which makes the proposed approach potentially useful for facilitating the development of integrated gasification design models. On the other hand, for soft-based modeling procedures can models successfully according to the data set and the given magnitudes' ranges. While in thermochemical, chemical and other physical processes, the experiments can be done with the magnitudes so different comparing to the training phase. Thus, the accuracy of modeling can decrease dramatically. But utilizing diverse main data training models may be costly. Limitations are always asserted besides the opportunities but, by developing efficient methods the capability of soft-based models can keep ahead of physical models by defining real world condition when programming AI oriented strategies.

References

[1] D. A. Akinpelu, O. A. Adekoya, P. O. Oladoye, C. C. Ogbaga, and J. A. Okolie, (2023) *"Machine learning applications in biomass pyrolysis: from biorefinery to end-of-life product management"* **Digital Chemical Engineering *8***: 100103. DOI: 10.1016/j.dche.2023.100103.

[2] S. Talatahari, M. Azizi, M. Tolouei, B. Talatahari, and P. Sareh, (2021) *"Crystal structure algorithm (CryStAl): a metaheuristic optimization method"* **IEEE Access *9***: 71244–71261. DOI: 10.1109/ACCESS.2021.3079161.

[3] T. Sağ, (2022) *"PVS: a new population-based vortex search algorithm with boosted exploration capability using polynomial mutation"* **Neural Computing and Applications *34***: 18211–18287. DOI: 10.1007/s00521-022-07671-x.

[4] A. Shafizadeh, H. Shahbeik, S. Rafiee, A. Moradi, M. Shahbaz, M. Madadi, C. Li, W. Peng, M. Tabatabaei, and M. Aghbashlo, (2023) *"Machine learning-based characterization of hydrochar from biomass: Implications for sustainable energy and material production"* **Fuel *347***: 128467. DOI: 10.1016/j.fuel.2023.128467.

[5] F. Elmaz, Ö. Yücel, and A. Y. Mutlu, (2020) *"Predictive modeling of biomass gasification with machine learning-based regression methods"* **Energy *191***: 116541. DOI: 10.1016/j.energy.2019.116541.

[6] H. Ullah, S. Khan, B. Chen, A. Shahab, L. Riaz, L. Lun, and N. Wu, (2023) *"Machine learning approach to predict adsorption capacity of Fe-modified biochar for selenium"* **Carbon Research 2**: 29. DOI: 10.1007/s44246-023-00061-5.

[7] M. S. Zaghloul, R. A. Hamza, O. T. Iorhemen, and J. H. Tay, (2020) *"Comparison of adaptive neuro-fuzzy inference systems (ANFIS) and support vector regression (SVR) for data-driven modelling of aerobic granular sludge reactors"* **Journal of Environmental Chemical Engineering *8***: 103742. DOI: 10.1016/j.jece.2020.103742.

[8] M. Awad, R. Khanna, M. Awad, and R. Khanna, (2015) *"Support vector regression"* **Efficient learning machines: Theories, concepts, and applications for engineers and system designers**: 67–80. DOI: 10.1007/978-1-4302-5990-9_4.

[9] J. Li, M. Suvarna, L. Pan, Y. Zhao, and X. Wang, (2021) *"A hybrid data-driven and mechanistic modelling approach for hydrothermal gasification"* **Applied Energy *304***: 117674. DOI: 10.1016/j.apenergy.2021.117674.

[10] S. Shenbagaraj, P. K. Sharma, A. K. Sharma, G. Raghav, K. B. Kota, and V. Ashokkumar, (2021) *"Gasification of food waste in supercritical water: An innovative synthesis gas composition prediction model based on Artificial Neural Networks"* **International Journal of Hydrogen Energy *46***: 12739–12757. DOI: 10.1016/j.ijhydene.2021.01.122.

[11] E. E. Ozbas, D. Aksu, A. Ongen, M. A. Aydin, and H. K. Ozcan, (2019) *"Hydrogen production via biomass gasification, and modeling by supervised machine learning algorithms"* **International Journal of Hydrogen Energy *44***: 17260–17268. DOI: 10.1016/j.ijhydene.2019.02.108.

[12] J. George, P. Arun, and C. Muraleedharan, (2018) *"Assessment of producer gas composition in air gasification of biomass using artificial neural network model"* **International Journal of Hydrogen Energy *43***: 9558–9568. DOI: 10.1016/j.ijhydene.2018.04.007.

[13] M. Ozonoh, B. O. Oboirien, A. Higginson, and M. O. Daramola, (2020) *"Performance evaluation of gasification system efficiency using artificial neural network"* **Renewable Energy *145***: 2253–2270. DOI: 10.1016/j.renene.2019.07.136.

[14] P. V. Gopirajan, K. P. Gopinath, G. Sivaranjani, and J. Arun, (2021) *"Optimization of hydrothermal gasification process through machine learning approach: Experimental conditions, product yield and pollution"* **Journal of Cleaner Production *306***: 127302. DOI: 10.1016/j.jclepro.2021.127302.

[15] A. Lee, P. Taylor, J. Kalpathy-Cramer, and A. Tufail, (2017) *"Machine learning has arrived!"* **Ophthalmology *124***: 1726–1728. DOI: 10.1016/j.ophtha.2017.08.046.

[16] S. L. Narnaware and N. L. Panwar, (2022) *"Biomass gasification for climate change mitigation and policy framework in India: A review"* **Bioresource Technology Reports *17***: 100892. DOI: 10.1016/j.biteb.2021.100892.

[17] Z. Ceylan and S. Ceylan. *Application of machine learning algorithms to predict the performance of coal gasification process*. Elsevier, 2021, 165–186. DOI: 10.1016/B978-0-12-821092-5.00003-6.

[18] S. Ascher, W. Sloan, I. Watson, and S. You, (2022) *"A comprehensive artificial neural network model for gasification process prediction"* **Applied Energy *320***: 119289. DOI: 10.1016/j.apenergy.2022.119289.

[19] Ö. Tezer, N. Karabağ, A. Öngen, C. Ö. Çolpan, and A. Ayol, (2022) *"Biomass gasification for sustainable energy production: A review"* **International Journal of Hydrogen Energy *47***: 15419–15433. DOI: 10.1016/j.ijhydene.2022.02.158.

[20] Y. Xiao, S. Xu, Y. Song, Y. Shan, C. Wang, and G. Wang, (2017) *"Biomass steam gasification for hydrogen-rich gas production in a decoupled dual loop gasification system"* **Fuel Processing Technology *165***: 54–61. DOI: 10.1016/j.fuproc.2017.05.013.

[21] D. Singh. *Advances in industrial waste management*. Elsevier, 2023, 385–416. DOI: 10.1016/B978-0-323-90463-6.00027-0.

[22] O. Ellabban, H. Abu-Rub, and F. Blaabjerg, (2014) *"Renewable energy resources: Current status, future prospects and their enabling technology"* **Renewable and sustainable energy reviews *39***: 748–764. DOI: 10.1016/j.rser.2014.07.113.

[23] D. Baruah, D. C. Baruah, and M. K. Hazarika, (2017) *"Artificial neural network based modeling of biomass gasification in fixed bed downdraft gasifiers"* **Biomass and Bioenergy *98***: 264–271. DOI: 10.1016/j.biombioe.2017.01.029.

[24] V. N. Vapnik and V. N. Vapnik, (2000) *"Controlling the generalization ability of learning processes"* **The Nature of Statistical Learning Theory**: 93–122. DOI: 10.1007/978-1-4757-3264-1_5.

[25] C. Loha, H. Chattopadhyay, and P. K. Chatterjee, (2013) *"Energy generation from fluidized bed gasification of rice husk"* **Journal of renewable and sustainable energy *5***: DOI: 10.1063/1.4816496.

[26] K. G. Mansaray, A. E. Ghaly, A. M. Al-Taweel, F. Hamdullahpur, and V. I. Ugursal, (1999) *"Air gasification of rice husk in a dual distributor type fluidized bed gasifier"* **Biomass and bioenergy *17***: 315–332. DOI: 10.1016/S0961-9534(99)00046-X.

[27] S. Sarker, F. Bimbela, J. L. Sánchez, and H. K. Nielsen, (2015) *"Characterization and pilot scale fluidized bed gasification of herbaceous biomass: A case study on alfalfa pellets"* **Energy Conversion and Management *91***: 451–458. DOI: 10.1016/j.enconman.2014.12.034.

[28] S. Sarker and H. K. Nielsen, (2015) *"Assessing the gasification potential of five woodchips species by employing a lab-scale fixed-bed downdraft reactor"* **Energy Conversion and Management *103***: 801–813. DOI: 10.1016/j.enconman.2015.07.022.

[29] S. Kaewluan and S. Pipatmanomai, (2011) *"Gasification of high moisture rubber woodchip with rubber waste in a bubbling fluidized bed"* **Fuel Processing Technology *92***: 671–677. DOI: 10.1016/j.fuproc.2010.11.026.

[30] J. M. D. Andres, A. Narros, and M. E. Rodríguez, (2011) *"Air-steam gasification of sewage sludge in a bubbling bed reactor: Effect of alumina as a primary catalyst"* **Fuel Processing Technology *92***: 433–440. DOI: 10.1016/j.fuproc.2010.10.006.

[31] P. M. Lv, Z. H. Xiong, J. Chang, C. Z. Wu, Y. Chen, and J. X. Zhu, (2004) *"An experimental study on biomass air–steam gasification in a fluidized bed"* **Bioresource technology *95***: 95–101. DOI: 10.1016/j.biortech.2004.02.003.

[32] P. Iovane, A. Donatelli, and A. Molino, (2013) *"Influence of feeding ratio on steam gasification of palm shells in a rotary kiln pilot plant. Experimental and numerical investigations"* **Biomass and Bioenergy *56***: 423–431. DOI: 10.1016/j.biombioe.2013.05.025.

[33] Z. Zhang and S. Pang, (2017) *"Experimental investigation of biomass devolatilization in steam gasification in a dual fluidised bed gasifier"* **Fuel *188***: 628–635. DOI: 10.1016/j.fuel.2016.10.074.

[34] S. Fremaux, S.-M. Beheshti, H. Ghassemi, and R. Shahsavan-Markadeh, (2015) *"An experimental study on hydrogen-rich gas production via steam gasification of biomass in a research-scale fluidized bed"* **Energy Conversion and Management *91***: 427–432. DOI: 10.1016/j.enconman.2014.12.048.

[35] J. Herguido, J. Corella, and J. Gonzalez-Saiz, (1992) *"Steam gasification of lignocellulosic residues in a fluidized bed at a small pilot scale. Effect of the type of feedstock"* **Industrial & engineering chemistry research *31***: 1274–1282.

[36] P. P. Dutta, V. Pandey, A. R. Das, S. Sen, and D. C. Baruah, (2014) *"Down draft gasification modelling and experimentation of some indigenous biomass for thermal applications"* **Energy Procedia *54***: 21–34. DOI: 10.1016/j.egypro.2014.07.246.

[37] E. Biagini, F. Barontini, and L. Tognotti, (2016) *"Development of a bi-equilibrium model for biomass gasification in a downdraft bed reactor"* **Bioresource technology *201***: 156–165. DOI: 10.1016/j.biortech.2015.11.057.

[38] T. Song, J. Wu, L. Shen, and J. Xiao, (2012) *"Experimental investigation on hydrogen production from biomass gasification in interconnected fluidized beds"* **Biomass and bioenergy *36***: 258–267. DOI: 10.1016/j.biombioe.2011.10.021.

[39] R. Yin, R. Liu, J. Wu, X. Wu, C. Sun, and C. Wu, (2012) *"Influence of particle size on performance of a pilot-scale fixed-bed gasification system"* **Bioresource technology *119***: 15–21. DOI: 10.1016/j.biortech.2012.05.085.

[40] I. Narvaez, A. Orio, M. P. Aznar, and J. Corella, (1996) *"Biomass gasification with air in an atmospheric bubbling fluidized bed. Effect of six operational variables on the quality of the produced raw gas"* **Industrial & Engineering Chemistry Research *35***: 2110–2120. DOI: 10.1021/ie9507540.

[41] P. Lahijani and Z. A. Zainal, (2011) *"Gasification of palm empty fruit bunch in a bubbling fluidized bed: a performance and agglomeration study"* **Bioresource technology *102***: 2068–2076. DOI: 10.1016/j.biortech.2010.09.101.

[42] C. Gai and Y. Dong, (2012) *"Experimental study on non-woody biomass gasification in a downdraft gasifier"* **International Journal of hydrogen energy *37***: 4935–4944. DOI: 10.1016/j.ijhydene.2011.12.031.

[43] H. Liu, J. Hu, H. Wang, C. Wang, and J. Li, (2012) *"Experimental studies of biomass gasification with air"* **Journal of natural gas chemistry *21***: 374–380. DOI: 10.1016/S1003-9953(11)60379-4.

[44] G. Ruoppolo, P. Ammendola, R. Chirone, and F. Miccio, (2012) *"H2-rich syngas production by fluidized bed gasification of biomass and plastic fuel"* **Waste Management *32***: 724–732. DOI: 10.1016/j.wasman.2011.12.004.

[45] A. Erkiaga, G. Lopez, M. Amutio, J. Bilbao, and M. Olazar, (2014) *"Influence of operating conditions on the steam gasification of biomass in a conical spouted bed reactor"* **Chemical engineering journal *237***: 259–267. DOI: 10.1016/j.cej.2013.10.018.

[46] C. V. Huynh and S.-C. Kong, (2013) *"Performance characteristics of a pilot-scale biomass gasifier using oxygen-enriched air and steam"* **Fuel *103***: 987–996. DOI: 10.1016/j.fuel.2012.09.033.

[47] H. Karatas and F. Akgun, (2018) *"Experimental results of gasification of walnut shell and pistachio shell in a bubbling fluidized bed gasifier under air and steam atmospheres"* **Fuel *214***: 285–292. DOI: 10.1016/j.fuel.2017.10.061.

[48] J. Wang, G. Cheng, Y. You, B. Xiao, S. Liu, P. He, D. Guo, X. Guo, and G. Zhang, (2012) *"Hydrogen-rich gas production by steam gasification of municipal solid waste (MSW) using NiO supported on modified dolomite"* **international journal of hydrogen energy *37***: 6503–6510. DOI: 10.1016/j.ijhydene.2012.01.070.

[49] Z. Khan, S. Yusup, M. M. Ahmad, and B. L. F. Chin, (2014) *"Hydrogen production from palm kernel shell via integrated catalytic adsorption (ICA) steam gasification"* **Energy Conversion and Management *87***: 1224–1230. DOI: 10.1016/j.enconman.2014.03.024.

[50] S. Luo, Y. Zhou, and C. Yi, (2012) *"Syngas production by catalytic steam gasification of municipal solid waste in fixed-bed reactor"* **Energy *44***: 391–395. DOI: 10.1016/j.energy.2012.06.016.

[51] M. Baratieri, E. Pieratti, T. Nordgreen, and M. Grigiante, (2010) *"Biomass gasification with dolomite as catalyst in a small fluidized bed experimental and modelling analysis"* **Waste and Biomass Valorization *1***: 283–291. DOI: 10.1007/s12649-010-9034-6.

[52] U. Arena, L. Zaccariello, and M. L. Mastellone, (2010) *"Fluidized bed gasification of waste-derived fuels"* **Waste Management *30***: 1212–1219. DOI: 10.1016/j.wasman.2010.01.038.

[53] U. Arena and F. D. Gregorio, (2014) *"Gasification of a solid recovered fuel in a pilot scale fluidized bed reactor"* **Fuel *117***: 528–536. DOI: 10.1016/j.fuel.2013.09.044.

[54] G. S. Margarat, S. Kumar, and S. Rajan. *Forecasting wind energy production using machine learning techniques*. 2023. DOI: 10.1051/e3sconf/202338701007.