Machine Learning Approach to Predict the Surface Charge Density of Monodispersed Particles in Gas–Solid Fluidized Beds

Junyu Lu,* Chenlong Duan, and Yuemin Zhao

ABSTRACT: Gas–solid fluidized beds are complex particle systems, and the electrostatic behavior of particles in fluidized beds is even more complex, which is influenced by numerous factors such as particle properties and operating conditions. Current studies focus on the effect of a certain factor on particle charging without a global picture. Furthermore, there is no mathematical model that can describe the interaction of multiple factors on particle charging because it is difficult to build a model for such a complex system. Therefore, a new approach is needed. In this study, a model capable of accurately predicting the surface charge density of particles in monodispersed gas–solid fluidized beds within a certain range was developed based on the literature and experimental data through several machine learning methods including kernel ridge regression (KRR), support vector machine regression (SVR), and multilayer perceptron (MLP). SVR and MLP models gave the best results with \( R^2 \) equal to 0.980 and 0.979, respectively. However, the sensitivity analysis showed that the MLP model was more reliable than the SVR model. In conclusion, the feasibility of using machine learning to analyze the charging behavior of particles in fluidized beds is demonstrated, and the proposed MLP model can serve as an accurate correlative tool for fast and effective estimation of particle surface charge density in gas–solid fluidized beds.

1. INTRODUCTION

Gas–solid fluidized beds are commonly used in chemical engineering industries as reactors for polymerization, catalytic cracking, gasification, separation, and combustion. During fluidization, electrostatic charges frequently develop due to triboelectric charging and gas ionization. The electrostatic charge may cause wall fouling, sheeting, and other issues. To control the electrostatic effect in fluidized beds, it is necessary to conduct a comprehensive study of the mechanism of particle charging in fluidized beds. However, particle charging in fluidized beds is closely related to many parameters such as the nature of the particles and the operating conditions. Currently, no equation can describe this complex system. Therefore, a new method linking the parameters is needed to solve this problem and to predict particle charging.

In the past few decades, researchers have conducted numerous experiments to investigate the effect of particle properties and operating conditions on the surface charge density (or charge-to-mass ratio). The surface charge density and the charge-to-mass ratio are both widely used to describe the degree of particle charging, and the two can be interconverted without loss. Generally, the charge-to-mass ratio can be obtained from two directly measurable values: total charge and total mass of particles. The surface charge density is suitable for comparing the degree of charging of different particles whose properties are known. However, because the actual surface area of particles is difficult to ascertain, the surface charge density is generally derived from equivalent area calculations.

Among the properties that influence the charge of a particle, the most important are its size, sphericity, and dielectric constant. The particle size is positively correlated with particle charging, implying that the degree of charging increases with increasing particle size. Saleh et al. proposed that particles having lower sphericities are more strongly charged after comparing crushed glass beads with the originals. The effect of the dielectric constant is complicated. McCarty and Whitesides reviewed the theory of contact electrification and proposed that a substrate having a higher dielectric constant can generally support a greater density of surface charge. However, the experimental results reported by Wu and Bi proposed different models.
show that the charge density of PE (lower dielectric constant) was higher than that of glass beads (higher dielectric constant.)

The operating conditions that influence particle charging are gas velocity, pressure, temperature, and relative humidity. Numerous studies have shown that the particle charge increases with the gas velocity in the bubbling or turbulent regimes because of the more turbulent particle motion at higher gas velocities. Moughrabiah et al. proposed that the degree of bed electrification increased with increasing absolute pressure and decreased with increasing temperature. The effect of humidity on particle charging is a complex topic for different systems; however, for dielectrics, increasing the relative humidity suppresses particle charging. Park et al. studied systems; however, for dielectrics, increasing the relative humidity because of the resulting enhancement in charge dissipation. However, the present studies have mainly focused on the effect of a change of one condition on the surface charge density without considering the joint effect of multiple operating conditions and particle properties.

Machine learning is a popular research tool for the statistical analysis of scientific data to solve problems that are difficult to solve using traditional methods. Machine learning allows predictions to be made without a theoretical basis and with a limited data set. Many studies have applied machine learning to model fluidization. Chew and Cococc applied the random forest method and neural networks to study the influence of factors on local mass flux and species segregation in circulating fluidized beds. Patel et al. used self-organizing map analysis to classify clusters in circulating fluidized bed risers containing Geldart Group B particles. Zhou et al. used text mining to obtain a vast amount of data on minimum fluidization velocities and the factors associated with them from the literature and used the extracted data to predict minimum fluidization velocities by employing artificial neural networks. Guo et al. used machine learning to enhance the electrical capacitance tomography measurement of gas–solid fluidized beds for online measurement. Grisar proposed a machine learning model that can predict the density of large molecules or condensed phases with a cost that scales linearly with the number of atoms.

On the other hand, machine learning is also widely used in industrial research. Rostami et al. applied support vector machines and multilayer neural networks and other methods for oil and gas processing and made accurate predictions for natural gas viscosity, viscoelastic surfactant viscosity, CO₂ absorption, and other critical parameters. And machine learning has a wide range of applications in industrial control such as natural gas processing, petroleum production, pharmaceutical manufacturing, and crystallization processes.

These studies show that machine learning has promising applications in the field of fluidization. However, there is an absence of models for predicting particle charging in fluidized beds, which is a complex, multivariate, and nonlinear problem. The feasibility of machine learning at predicting the surface charge density of particles in fluidized beds needs to be investigated.

The goal of this study is to establish a model for accurately predicting the surface charge density of particles in fluidized beds based on previously reported and currently generated data. To achieve this objective, a data set containing two operating conditions (gas velocity and relative humidity) and three particle properties (particle size, minimum fluidization velocity, and relative permittivity) was constructed. The data set was then subjected to statistical analysis to understand the effect of the variables on the results and to minimize the interaction of the variables in Section 2. Furthermore, five methods (multivariate linear, polynomial, kernel ridge regression, support vector machine, and multilayer perceptron) were used to model and cross-validate the data set and compare their accuracy in Section 3. Then, in Section 4, the MLP and SVR models were then subjected to sensitivity analysis to understand the contribution of each input variable to the output. This study demonstrated the feasibility of using machine learning to analyze the charging behavior of particles in fluidized beds, and the proposed MLP model can serve as an accurate correlating tool for fast and effective estimation of particle surface charge density in gas–solid fluidized beds.

2. MATERIALS AND METHODS

2.1. Experimental Setup. Figure 1 schematically illustrates the experimental setup. For the fluidized bed, an acrylic column (height, h_{bed} = 600 mm; inner diameter, d_{out} = 100 mm) was placed on a metal gas predistribution chamber, and a double-layer cloth was sandwiched between the two as a gas distributor. The air was pumped in by a Roots blower, and subsequently flowed through a gas tank, an air dryer, and a rotary flow meter, and finally into the fluidized bed. The air dryer was an acrylic tube filled with color-changing silica gel and molecular sieves to maintain the relative humidity at ~20%. Particles were fluidized for 15 min to attain charge saturation. The charge of the particles was measured using a Faraday cup connected to an electrometer (Keithley 6514).

2.2. Construction of Data Sets. The original data set consisted of 61 points of data obtained from the literature and 25 points of data obtained from experiments performed during this work. After removing the outliers using the local outlier factor (LOF) method, 5 data points were removed and 81 remained. The properties of three types of particles (porcelain, coal powder, and glass beads), as determined experimentally or adapted from the literature, are summarized in Table 1.
Table 1. Particle Properties

| Source      | Material     | Particle size $d_p$ (μm) | $u_{ref}$ (m/s) | Relative permittivity $\varepsilon_r$ |
|-------------|--------------|--------------------------|-----------------|--------------------------------------|
| Tardos 1980 | porcelain    | 2000                     | 0.98            | 6                                    |
| Murtonmaa 2003 | glass beads | 1130                     | 1.5             | 3                                    |
| He 2015     | glass beads  | 624                      | 0.27            | 3                                    |
| experiment  | coal         | 144                      | 0.064           | 10                                   |
| experiment  | coal         | 487                      | 0.028           | 10                                   |
| experiment  | glass beads  | 243                      | 0.022           | 3                                    |

The data set contains five independent variables: the particle size, $d_p$, the relative permittivity (also known as dielectric constant) of the particle, $\varepsilon_r$; the relative humidity, RH; the minimum fluidization velocity, $u_{mf}$, and the superficial gas velocity, $u_f$. Fotovat et al. summarized the factors influencing electrostatics in fluidized beds; after eliminating the factors that have negligible influence on the results, four factors representing the particle properties and operating conditions, respectively, were selected. Gas velocity $u_f$ is one of the most important operating conditions in fluidized beds and influences the movement of the particles within the bed. The relative humidity RH has an important effect on the charge of particles in fluidized beds, and it has been reported in the literature that the particle charge decreases with increasing relative humidity. The particle size and the dielectric constant both influence the electrostatic properties of the particles. The particles modeled and used in this study were all monodispersed.

The response variable in this data set is the surface charge density, $\sigma$. The charge-to-mass ratio is more frequently used in other studies to indicate the degree of particle charging; however, this variable can vary considerably when studying particles of different sizes. Furthermore, for nonconducting particles, the charge will accumulate on the surface of the particle rather than in its bulk. Therefore, in this study, the charge-to-mass ratio was converted into an equivalent surface charge density according to eq 1 to accommodate different particle sizes. Figure 2 shows the histograms for the variables, the distribution of the data is uneven and highly discrete. Among all of the input variables, only $u_f$ is continuously changing in two ranges. This is because $u_{mf}$, $d_p$, RH, and $\varepsilon_r$ are fixed experimental conditions, both in the literature and in this study, and the range of $u_f$ is determined by these conditions.

\[
\sigma = \frac{q_p \rho d_p}{6m} = \frac{q_m \rho d_p}{6}
\]  

(1)

2.3. Modeling Procedure. Figure 3 schematically illustrates the steps of the modeling process. Outlier detection was first performed on the raw data obtained from the literature and experiments to remove outliers. The cleaned data were then statistically analyzed to study the significance and dominance of each variable. At the same time, the multicollinearity between variables was analyzed, and if it was high, it was eliminated by merging the variables in a reasonable way. Some methods (such as MLP) require regularized data; hence, the data are regularized for these methods. The preprocessed data were then divided into training data and testing data. The model was trained with the training data and then cross-validated with the test data. The hyperparameters were adjusted according to the validation results, and the model was trained again until the accuracy of the model was high enough. Finally, a sensitivity analysis was performed on the model to confirm whether its response to the variables was reasonable.

2.4. Validation of Models. 2.4.1. n-Fold Cross-Validation. In machine learning, the data are generally divided into a training set and a test set. A randomly selected portion (generally 80%) of the original data set is used to train the model, while the remaining portion is used to test the validity of the model. The reason for not using all of the data points for training and testing is to avoid overfitting, which refers to a training error wherein the model corresponds excessively closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably. For small data sets, n-fold cross-validation is frequently used to maximize the use of the data. The principle of n-fold cross-validation is to randomly select and divide the data set into n parts and cycle the use of one of the parts as the test set and the others as the training set. The advantage of this method is that it allows all data points to be rotated as training and test sets, thereby maximizing data usage. In this study, a 4-fold cross-validation was applied for training and testing.

2.4.2. Coefficient of Determination. The coefficient of determination, denoted as $R^2$, is a measure of the proximity of the value predicted by the model to the observed value. The definition of the coefficient of determination is

\[
R^2 = 1 - \frac{SS_{res}}{SS_{total}}
\]  

(2)

where $SS_{res}$ is the sum of the squares of the residuals

\[
SS_{res} = \sum_i (y_i - f_i)^2
\]  

(3)

$y_i$ is the observed value, and $f_i$ is the predicted value. $SS_{total}$ is the total sum of squares

\[
SS_{total} = \sum_i (y_i - \bar{y})^2
\]  

(4)

and $\bar{y}$ is the mean of the observed data.
Generally, the value of $R^2$ lies between 0 and 1, and values closer to 1 indicate accurate prediction by the model. In case of excessive errors in prediction, the value of $R^2$ will be lower than 0, which indicates that the model is completely unreliable.

2.4.3. Root-Mean-Square Error. The root-mean-square error (RMSE)\(^{46}\) is the distance between the predicted and observed values, and is defined as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - f_i)^2}{n}}$$

(5)

It differs from $R^2$ in that $R^2$ reflects the level of confidence level of the model, whereas RMSE reflects the absolute error in the predictions made by the model. The mean squared error is also used in some other practices; however, the dimension of the RMSE is identical to that of the observed values and can be substituted into units to evaluate the significance of the error.

2.5. Multiple and Polynomial Regression with the Least-Squares Method. Linear regression is the most fundamental method for analyzing the relationship between the response variable and the independent variables. In the case of multiple independent variables, as in this study, the linear regression can be called multiple linear regression (MLR). The concept of multiple linear regression is to assume an equation, such as eq 6, and manipulate the parameters to achieve the closest fit. To optimize the parameters, the least-squares method\(^{47}\) was implemented. The independent variables $x$, response variable $y$, and parameters $\beta$ are written as vectors $X$, $y$, and $\beta$ ($X$ is a matrix because there are multiple independent variables). The loss function can thereafter be written as eq 7. Since the loss function is convex, the optimal values of the parameters are obtained when its gradient is zero, as described in eqs 8 and 9.

\begin{equation}
y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4
\end{equation}

(6)

\begin{equation}
L(\hat{\beta}) = \| X \hat{\beta} - y \|^2
\end{equation}

(7)

\begin{equation}
\frac{\partial L(\hat{\beta})}{\partial \hat{\beta}} = -2X^T \hat{y} + 2X^T X \hat{\beta} = 0
\end{equation}

(8)

\begin{equation}
\hat{\beta} = (X^T X)^{-1} X^T y
\end{equation}

(9)

The equation of the multiple linear regression is a plane in a multidimensional space; however, the equation relating the factors with the surface charge density may be a hyperplane. Therefore, the equation of this surface can be assumed to be a quadratic, cubic, or higher-order polynomial, as described in eq 10. Although the hyperplane is nonlinear, statistically, it belongs to a particular type of MLR. This method is also known as polynomial regression (PR).

\begin{equation}
y = \alpha + \beta_1 x_1 + \cdots + \beta_2 x_2 x_2 + \cdots + \beta_4 x_4^2 \quad \text{degree} = 2
\end{equation}

(10)

Table 2 shows the $R^2$ and RMSE of the polynomial regression with different degrees. At degrees exceeding 3, the $R^2$ of the model corresponding to the test set is lower than 0, indicating that overfitting has occurred at these degrees and the models are invalid. Therefore, the degree of the polynomial was set to 2, whereupon the regression equation became a hyperplane in multidimensional space. When the degree is 1, the model is an MLR equation.
Table 2. $R^2$ and RMSE of the Polynomial Regressions of Different Degrees

| degree | $R^2$ | RMSE | standard deviation of $R^2$ | standard deviation of RMSE |
|--------|-------|------|-----------------------------|---------------------------|
| 1      | 0.727 | 0.307| 0.069                       | 0.036                     |
| 2      | 0.932 | 0.148| 0.022                       | 0.033                     |
| 3      | -20.173 | 1.786 | 31.850                     | 2.206                     |
| 4      | -7.6 x 10$^6$ | 687.055 | 1.3 x 10$^7$      | 1115.662                  |
| 5      | -2.7 x 10$^6$ | 76.552 | 3.4 x 10$^4$               | 64.921                    |

2.6. Kernel Ridge Regression. Ridge regression, as known as Tikhonov regularization, is not an independent regression method but a regularization method, which is an improved form of linear regression. Ordinary linear regression may produce a large bias when the amount of data is small or when there are outliers. To prevent some parameters from being overweighed, a penalty $\lambda I$ is introduced in the equation of the parameters $\hat{\beta}$ (eq 9), as shown in eq 11.

$$\hat{\beta}_R = (X^T X + \lambda I)^{-1}X^T y$$

(eq 11)

For the reasons mentioned previously, the correspondence between the factors and the surface charge density may be a nonlinear model existing on a high-dimensional space. Although it is possible to map the data into high-dimensional spaces through feature expansion, it is simpler to do so through a kernel trick: using a matrix inversion lemma, eq 11 can be rewritten as

$$\hat{\beta}_R = X^T(\lambda I + XX^T)^{-1}y$$

(eq 12)

let a function $K$ replace $XX^T$, then

$$\hat{\beta}_R = X^T(\lambda I + K)^{-1}y$$

(eq 13)

This is known as the kernel ridge regression (KRR), and the function $K$ is known as the kernel function. Different kernel functions can be applied to KRR to determine the most suitable one.

Five kernel functions—linear, radial basis function (RBF), sigmoid, polynomial (poly), and Laplace kernels—were applied with 4-fold cross-validation, and their $R^2$ and RMSE were calculated and are shown in Figure 4. The $R^2$ of the sigmoid kernel was lower than 0; therefore, it is unsuitable for this data set. The remaining four kernels made valid predictions, and the $R^2$ of the RBF, polynomial, and Laplacian kernels were greater than 0.9. Consequently, these three functions were selected as the kernels.

The polynomial kernel was similar to the polynomial features as in Section 3.3, except that there was no need to transform the input space to higher dimensions while using the kernel trick. Therefore, the appropriate degree of the polynomial kernel was 2. The radial basis function, as shown in eq 14, is a kernel function based on the Euclidean distance between data points. Its feature space has an infinite number of dimensions. Therefore, the RBF kernel can determine the correlations between the data in high-dimensional spaces. The Laplace kernel, described in eq 15, is a variant of the RBF kernel, which differs in that the Manhattan distance is used instead of the Euclidean distance.

$$K_{RBF} = \exp(-\gamma \|x - x'\|_2^2)$$

(eq 14)

$$K_{Laplacian} = \exp(-\gamma \|x - x'\|_1)$$

(eq 15)

The hyperparameter $\gamma$ is a correction parameter for distance in the kernel method that influences the confidence of the model. Figure 5 shows the $R^2$ for the RBF and Laplacian kernels as a function of $\gamma$, wherein the range of $\gamma$ is 0.001–100, and the $\gamma$ corresponding to the maximum $R^2$ is optimal. Therefore, the $\gamma$ of the RBF and Laplacian kernels was set to 3 and 1, respectively, which creates a negligible difference.

2.7. Support Vector Regression Machine. The support vector regression machine (SVR) proposed by Vapnik is a regression method based on an $\epsilon$-insensitive loss function. The fundamental principle of SVR is to identify the hyperplane that maximizes the margin (distance from data points to prediction function) while a part of the error is tolerated.

Assuming a function such as $f(x) = (w,x) + b$, where $w$ and $b$ are parameter factors, and setting a variable $\epsilon$ to indicate the width of the tube, where the points in the tube will be ignored, this problem can be written as a convex optimization problem

$$\min \frac{1}{2} \|w\|^2$$

subject to

$$\chi - f(x_i) \leq \epsilon$$

$$f(x_i) - \chi \leq \epsilon$$

(eq 16)

However, the actual data set may contain points that are exceptionally distant from the others, which may lead to large
deviations. Therefore, the slack variables ξ_i and ξ_i* are introduced to avoid it

\[
\begin{align*}
\min & \quad \frac{1}{2} \|w\|^2 + C \sum (\xi_i + \xi_i^*) \\
\text{s. t.} & \quad y_i - f(x_i) \leq \epsilon + \xi_i \\
& \quad f(x_i) - y_i \leq \epsilon + \xi_i^* \\
& \quad \xi_i, \xi_i^* \geq 0
\end{align*}
\]  
(17)

Since the process of finding the solution of SVR is complicated and not the focus of this study, the derivation is omitted. The solution of SVR is described in eq 18

\[
f(x) = \sum (\alpha_i - \alpha_i^*)K(x, x) + b
\]

where α_i and α_i* are Lagrange multipliers. Similar to KRR, the kernel trick was applied to perform regressions on higher-dimensional spaces, and eq 18 can be rewritten as

\[
f(x) = \sum (\alpha_i - \alpha_i^*)K(x, x) + b
\]

Since the data set in this study is finitely indistinguishable in low dimensions, the RBF kernel (eq 14) was selected as the kernel function.

In the SVR equations, there are two hyperparameters that need to be determined:\textsuperscript{55} γ in eq 14, which influences the kernel function, and C in eq 17, which influences the error tolerance. A grid search was conducted to identify the optimal combination of γ and C, and the results are shown in Figure 6.

Figure 6. Contour plot for γ and C of 4-fold cross-validation R^2 for SVR with RBF kernel (the dots indicate the optimal combinations; the red dot was selected; C = 100, γ = 0.63).

The blue area represents that the model gave unacceptable accuracy with this combination of parameters, and the opposite is true for the red color, where the darker the red color represents the higher the accuracy of the model. The dots indicate the top 10 highest accuracies, and the red dots indicate the best accuracy, where the combination of parameters is optimal. The optimal combination, C = 100, γ = 0.63, which exhibited the highest R^2, was selected as the hyperparameter for SVR.

2.8. Multilayer Perceptron Regression. The multilayer perceptron (MLP)\textsuperscript{56,57} is a class of feedforward artificial neural network (ANN). It consists of the following three parts: one input layer, a minimum of one hidden layer, and one output layer. Figure 7 shows the multilayer perceptron network used in this study. The input layer contained four input variables, corresponding to the four independent variables of the data set. Since this MLP was used to solve the regression problem, the output layer had a single output variable, corresponding to the predicted values. Selecting the number of hidden layers and the number of neurons in each layer is a complex problem. In theory, a sufficiently large number of layers and neurons can solve any problem. However, an excessively large number of layers and neurons can lead to overfitting and severely retard the computation. Although there is no theoretical method to determine the number of layers and nodes, there are some empirical methods that can help. In this study, a method named pruning was employed.\textsuperscript{58} We initially assumed sufficient layers and nodes (i.e., 3 layers, and 10 nodes for each layer), and subsequently reduced the number in a stepwise manner until the minimum number required to obtain satisfactory results was identified (2 layers, 6 and 5 nodes for the first and second hidden layers, respectively).

The fundamental principle of MLP can be described as

\[
h_i = b_i + \sum \phi(w_i x_i)
\]

where h is a neuron in a hidden layer (i.e., \(h_{1,1}\)), x is the value from the preceding layer (i.e., \(x_i\)), w is the weight, and b is the bias. Since the data distribution may be nonlinear and the weighted input \(wx\) is clearly a linear function, an activation function \(\phi(x)\) was introduced to perform the nonlinear transformation. The principle of the activation function is similar to that of the aforementioned kernel trick, except the activation function also compresses the value to a restricted range. The activation function used in this study is the hyperbolic tangent function,\textsuperscript{59} which is described in eq 21, and the range of \(\phi(x)\) is \((-1,1)\).

\[
\phi(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}
\]

(21)

Therefore, the MLP network containing two hidden layers can be denoted as
\[ a_1 = b_i + \sum \phi(w_{ij} h_{2,i}) \]
\[ = b_i + \sum \phi\left(w_{ij}\phi\left(b_{2,i} + \sum \phi(w_{ik} h_{1,k})\right)\right) \]
\[ = b_i + \sum \phi\left(w_{ij}\phi\left(b_{2,i} + \sum \phi\left(w_{ik}\phi\left(b_{1,k} + \sum w_{ik} r_i\right)\right)\right)\right) \]  

(22)

Since the input variables are weighted and transformed by the activation function and passed to the neurons in the next layer, this structure is known as the feedforward structure.

The training process of MLP involves continuously adjusting the values of the weights \((w)\) and biases \((b)\) to minimize the deviation \((E)\)

\[ E = \frac{1}{2} \sum_{i=1}^{2} (y_i - o_i) \]  

(23)

where \(y_i\)’s are the observed values. However, even in the relatively simple MLP, numerous \(w\) and \(b\) values need to be determined. Determining them by traversal is extremely time- and resource-intensive. Therefore, the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) method\(^{[9]}\) was employed to estimate the values of the parameters. L-BFGS is a complex optimization algorithm based on quasi-Newton methods that does not contain hyperparameters that influence the results. Therefore, it is not discussed in detail here because it is not the focus of this paper. Compared to other optimizers, such as stochastic gradient descent, L-BFGS is more suitable for smaller data sets and delivers superior performance.

In contrast to other methods, although MLP is a weighted regression method, it remains sensitive to the scale of the input variables. To prevent inaccurate fitting due to different scales of variables, feature scaling of the data is required. Z-score normalization was performed to rescale the data

\[ x' = \frac{x - \bar{x}}{\sigma} \]  

(24)

where \(\sigma\) is the standard deviation. The z-score normalization altered the mean and variance of the data to 0 and 1, respectively, but did not alter the distribution.

3. RESULTS AND DISCUSSION

3.1. Statistical Analysis. To understand the effect of the variables on the surface charge density, the Pearson correlation coefficient is generally used for statistical analysis. To avoid the influence of other variables, the analysis is usually performed using the partial correlation coefficient, which controls for the correlation of other variables. Table 3 shows the partial correlation coefficient of variables with surface charge density, where the correlation coefficient \(r\) indicates the extent of influence that a particular variable has on the corresponding variable, and the \(p\)-value indicates the probability that it is not linearly correlated. In practice, a linear correlation can be considered when the \(p\)-value is lower than 0.05. As reported in Table 3, the \(p\)-values of all of the variables are lower than 0.05, indicating that they are linearly correlated with the surface charge density. The variables \(u_g\), \(u_{hub}\), \(d_p\), and \(e_i\) were positively correlated with the surface charge density, which increased with these variables. Conversely, RH was negatively correlated with the surface charge density, which decreased with increasing RH. This conclusion is consistent with the conclusions summarized by Fotovat et al.\(^{[12]}\) from different sources.

Although the partial correlation analysis explains the magnitude of the effect of each variable on the degree of change in the target, it does not explain the degree of contribution of each variable to the change in the target. To understand and compare the importance of each variable to the prediction model, a dominance analysis\(^{[61,62]}\) was conducted. The fundamental principle of the dominance analysis is to construct a model matrix containing every possible combination of each variable and find the \(R^2\) of each model. For example, the combinations of three variables \(x_1, x_2,\) and \(x_3\) are \(x_1, x_2, x_3, x_1 x_2, x_1 x_3, x_2 x_3,\) and \(x_1 x_2 x_3\); therefore, the dominance of \(x_1\) can be written as

\[ D_{x_1} = \begin{cases} R^2_{y|x_1} & k = 0 \\ \frac{1}{2}\left( (R^2_{y|x_1} - R^2_{y|x_2}) + (R^2_{y|x_1} - R^2_{y|x_3}) \right) & k = 1 \\ R^2_{y|x_1|x_2} - R^2_{y|x_3} & k = 2 \end{cases} \]  

(25)

where \(D\) is the dominance of a particular variable, and a higher \(D\) indicates that that particular variable is more important than the others. \(k\) is the size of the subset model: for \(k = 0\), \(D^{(0)}\) is addressed as the individual dominance, which can be interpreted as the variability explained solely by that particular variable. For \(k = n - 1\), \(D^{(n)}\) (in this case) is the interactional dominance and indicates the impact of a particular variable in the presence of all other predictors. For values of \(k\) between 0 and \(n - 1\), \(D^{(k)}\) (in this case) is the partial dominance and indicates the impact of a particular variable in all possible combinations with other variables. Especially, when \(n > 3\), \(D^{(1 < n < 2)}\) is the average of all of the partial dominances and should be addressed as the average partial dominance. To summarize the different types of dominance, the total dominance \(D_{total}\) can be calculated by averaging all of the conditional values, and it represents the extent to which a particular variable contributes to the \(R^2\) of the entire model. In other words, the sum of the \(D_{total}\) of all variables is exactly equal to the \(R^2\) of the entire model.

To analyze the dominance of the five variables in this study, a model matrix containing a total of \(2^5-1\) models was built, and four types of dominance were calculated for each variable, as listed in Table 4. The total dominance of the entire model was 0.8113, and the relative importance, which is the shear of the total dominance, for each variable, was calculated and summarized in Table 4. The RH and \(e_i\) have the highest relative importance, which indicates that these two variables most strongly influence the surface charge density in the fluidized bed. Additionally, this result indicates that the RH and the particle properties \(d_p\) and \(e_i\) have a stronger effect on the surface charge density than does the gas velocity.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
variable & correlation coefficient \(r\) & \(p\)-value \\
\hline
\(u_g\) & 0.5322 & \(3.57 \times 10^{-6}\) \\
\(u_{hub}\) & 0.5137 & \(8.75 \times 10^{-6}\) \\
\(d_p\) & 0.5970 & \(9.72 \times 10^{-9}\) \\
RH & -0.6722 & \(4.69 \times 10^{-10}\) \\
e_i & 0.6544 & \(1.90 \times 10^{-9}\) \\
\hline
\end{tabular}
\caption{Partial Pearson Correlation Coefficient of Variables with Surface Charge Density}
\end{table}
Since there are multiple independent variables in the data set, multicollinearity needs to be discussed. Multicollinearity is a phenomenon in which one of the independent variables can be linearly predicted from the others with a substantial degree of accuracy. In this situation, the coefficient estimates of the regression models may change significantly in response to small changes in the data. To test the collinearity of the variables, the variance inflation factors (VIFs) were calculated for each variable, as reported in Table 4. In practice, VIF > 10 indicates a strong collinearity of the variable. As reported in Table 4, the VIFs for \( d_p, u, u_{mf} \) and \( \varepsilon_c \) were close to 1, which means these three variables have significant covariance with each other. The reason behind this phenomenon is that a larger particle size corresponds to a higher \( u_{mf} \) and \( u \) needs to be higher than \( u_{mf} \) to achieve fluidization. Therefore, there is a strong logical relationship between these three variables.

Although multicollinearity has a smaller impact on some models (e.g., SVR and MLP), it needs to be addressed. Additionally, the states of beds containing particles having different properties at identical gas velocity may vary greatly, and cannot be compared directly. Therefore, \( u \) and \( u_{mf} \) were combined into the fluidization number \( u_f/u_{mf} \) which is commonly used in fluidization literature. The VIFs were calculated and the correlation matrix of variables with the fluidization number was formed as shown in Table 6. The results show that the covariance between variables is resolved. Therefore, the fluidization number \( u_f/u_{mf} \) is used to replace \( u \) and \( u_{mf} \) in the training and testing of the models.

### Table 5. Correlation Matrix and Variance Inflation Factors for Independent Variables

| variable | \( u \) | \( u_{mf} \) | \( d_p \) | RH | \( \varepsilon_c \) |
|----------|-------|-------|-------|----|---------|
| \( u \)  | 0.9445|       |       |     |         |
| \( u_{mf} \) | 0.9387| 0.8834|       |     |         |
| \( d_p \) | -0.0391| -0.0002| -0.0957|    |         |
| RH      | 0.1386| 0.0374| 0.2043| -0.6238| |
| \( \varepsilon_c \) | 54.27 | 27.03 | 31.05 | 3.56 | 4.22 |

### Table 6. Correlation Matrix and Variance Inflation Factors for Independent Variables (with Fluidization Number)

| \( d_p \) | \( u_f/u_{mf} \) | RH | \( \varepsilon_c \) |
|-------|---------------|----|---------|
| \( d_p \) | -0.3565 |       |           |
| RH    | -0.3374 | -0.0957|           |
| \( \varepsilon_c \) | 0.5667 | 0.2043| -0.6238 |
| VIF   | 2.71     | 5.60 | 3.53    | 8.93 |

### 3.2. Comparative Evaluation of Model Performance.

The seven regression models proposed (including KRR with three different kernel functions) were trained using four input variables (\( d_p, u, u_{mf}, \) RH, and \( \varepsilon_c \)) and employed to develop predictions for a single output variable (\( \sigma \)). The parameters of the proposed models are summarized in Table 7.

### Table 7. Parameters for the Proposed Models

| model | type | method | kernel (activation) function | parameters |
|-------|------|--------|------------------------------|------------|
| linear regression | linear | least squares | degree = 2 |
| polynomial regression | nonlinear | least squares | polynomial | degree = 2 |
| KRR | nonlinear | ridge regression | RBF | \( \gamma = 3 \) |
| SVR | nonlinear | support vector machine | RBF | \( \gamma = 0.63, C = 100 \) |
| MLP | nonlinear | L-BFGS | hyperbolic tangent | 2 layers, 6 and 5 neurons each layer |

A 4-fold cross-validation using \( R^2 \) and RMSE was performed for all of the prediction models, as shown in Figure 8, to compare their performance. To obtain a clearer understanding of the performances of the models, the predicted values of the output variable were plotted against the observed values, as shown in Figure 9. The reference line, whose slope is equal to 1, indicates the ideal results: the predicted values being exactly equal to the observed values. Therefore, the closer the points on the plots are to the reference line, the more accurate is the prediction.
The performance of the linear model was the lowest among all of the models; its $R^2$ and RMSE were 0.732 and 0.303, respectively. The distribution of its results, although in the vicinity of the ideal results, deviates significantly. The linear model could predict the direction of the surface charge density variation with variations in the parameters but was unable to make accurate predictions.

Compared to the linear model, the performance of the polynomial model with feature expansion was significantly higher. Its $R^2$ and RMSE were 0.922 and 0.164, respectively. As a different approach to the same idea, the kernel ridge regression with the polynomial kernel (KRR polynomial) yielded similar results ($R^2 = 0.915$, RMSE = 0.170.) The distribution of their results was also identical, with the majority of predictions being close to the ideal value and a minority of predictions deviating. When the degree was 2, both models could accurately predict the trend of surface charge density variation, and the predicted values could be used as references.

The kernel ridge regression with the RBF and Laplacian kernels (KRR_RBF and KRR_Laplacian) are an almost identical pair of models and yielded identical results ($R^2 = 0.956$, RMSE = 0.124). Both models performed well on the majority of the data points but exhibited biases on some discrete data points. This indicates that they are biased toward the majority of the results and ignore a small number of points. This is not a bad sign because these points may be outliers; however, this is uncertain in the absence of sufficient evidence.

Support vector machine (SVR) and multilayer perceptron regression (MLP), two popular machine learning methods, achieved the most accurate results ($R^2_{SVR} = 0.980$, RMSE$_{SVR} = 0.084$, $R^2_{MLP} = 0.979$, and RMSE$_{MLP} = 0.086$.) For the data set used in this study, both methods could accurately and reliably predict the variation in surface charge density.

### 3.3. Sensitivity Analysis

Since the models obtained with SVR and MLP are implicit models, also known as black-box models, their relationships between inputs and outputs may be poorly understood. To evaluate how much each input is contributing to the output uncertainty, a sensitivity analysis was performed. There are numerous methods to perform sensitivity analysis, mainly divided into local sensitivity analysis and global sensitivity analysis. Local sensitivity analysis, i.e., OAT (one-at-a-time) method, is easy to implement but usually ignores the interaction between inputs. Considering the possible interaction between the inputs in this study, it is preferable to choose the global sensitivity analysis method. The Sobola method, referred to as the variance-based sensitivity analysis, which is one of the most commonly used methods to perform global sensitivity analysis, was chosen to perform with the models. First, a matrix (size of 1024 × (2 × 4 + 2)) was generated with Saltelli’s sampling method as the sample matrix. $X$ was then input into the MLP model to obtain the output $Y$. Then the first-order sensitivity index $S_i$, which represents the contribution to the output variance of each input parameter alone averaged over variations in other input parameters, can be calculated with eq 26. And the total-order index $S_{Ti}$, which measures the contribution to the output variance of each input parameter including all variance caused by its interactions, can be calculated with eq 27.

\[
S_i = \frac{\text{Var}_i (E_{X_i} (Y | X_i))}{\text{Var}(Y)}
\]

\[
S_{Ti} = \frac{E_{X_i} (\text{Var}_i (Y | X_i))}{\text{Var}(Y)}
\]

Figure 10 shows the first-order and total-order sensitivity indices for the MLP model. For all input variables, their total-order sensitivity indices were significantly higher than their first-order indices. This indicates that for the MLP model, the contribution of the interaction of input variables on the high-dimensional space was greater than that of the individual variables. In the total-order sensitivity indices, the index of relative humidity is the highest (0.793), the index of particle size is next to the highest (0.565), and the fluidization number and dielectric constant are lowest (0.273 and 0.199).
The distribution of the sensitivity indices of each input variable is the same as the dominance analysis of the original data. This indicates that the MLP model can reflect the influence of each input variable on the output variable.

Figure 11 shows the first-order and total-order sensitivity indices for the SVR model. The total-order sensitivity indices were also significantly higher than the first-order indices, which indicates that the contribution of the interaction of input variables on the was greater than that of the individual variables. However, the sensitivity indices of the SVR model are distributed in a different way than those of the MLP model. The output responds most strongly to the fluidization number, followed by the particle size and dielectric constant. The relative humidity has the lowest sensitivity index, while it should be the most influential parameter. This may be due to the fact that $d_p$, RH, and $\varepsilon_r$ in the input data are phase-varying, while $u/u_{mf}$ is continuously varying, resulting in hard-to-detect overfitting of the SVR model.

Comparing these two models, the relationship between input variables and output in the MLP model is much closer to the real situation. In contrast, the SVR can only reflect the continuously varying fluidization number versus output under conditions close to the training data. Therefore, MLP is the most trustworthy predictor among these models.

4. CONCLUSIONS

In this study, several prediction models were built and trained using several machine learning methods including kernel ridge regression (KRR), support vector machine regression (SVR), and multilayer perceptron (MLP). The accuracy of the proposed models was investigated through cross-validation. The results show that the SVR and MLP models gave the best accuracy ($R_{SVR}^2 = 0.980$ and $R_{MLP}^2 = 0.979$). Among these two models, the sensitivity analysis shows that the MLP model reflects the relationship between input variables and output variables more realistically than the SVR model and agrees with the statistical analysis of the original data. Therefore, it can be stated that the MLP can be considered as an accurate predictor for particle surface charge density in gas–solid beds.

However, despite its ability to provide accurate predictions, the MLP model still has drawbacks. The models obtained with MLP (even with SVR and KRR) are implicit functions, which means it is difficult to describe the relationship between input variables and outputs with mathematical equations. Furthermore, it is difficult to verify the extensibility of the models because the range of the data set used in this study is narrow.

In conclusion, the proposed MLP model can serve as a fast and effective predictor for particle surface charge density in gas–solid fluidized beds. This model will help researchers and engineers find the charge density of particles within a fluidized bed without extensive testing and help them determine if relevant actions are needed. In future work, more data need to be added to improve the accuracy and extend the range of these models. Furthermore, many fluidized bed practices use binary or multiple mixed particles, and the same method can be used to analyze multiple mixed particles to study the mechanism of multiple particle interactions.

## Author Information

### Corresponding Author

Junyu Lu — School of Chemical Engineering and Technology, China University of Mining and Technology, Xuzhou 221116, China; Department of Chemical Engineering, Kyoto University, Kyoto 615-8510, Japan; orcid.org/0000-0002-9213-8976; Email: mailolifjy@gmail.com

### Authors

Chenlong Duan — School of Chemical Engineering and Technology, China University of Mining and Technology, Xuzhou 221116, China; orcid.org/0000-0002-8093-6719

Yuemin Zhao — School of Chemical Engineering and Technology, China University of Mining and Technology, Xuzhou 221116, China

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c00299

### Notes

The authors declare no competing financial interest.

### Acknowledgments

This work was supported by the joint Ph.D. program of the “double first rate” construction disciplines of the China University of Mining & Technology (CUMT).

### References

1. Chatzidoukas, C.; Perkins, J. D.; Pistikopoulos, E. N.; Kiparissides, C. Optimal Grade Transition and Selection of Closed-Loop Controllers in a Gas-Phase Olefin Polymerization Fluidized Bed Reactor. Chem. Eng. Sci. 2003, 58, 3643–3658.

2. Rodríguez, E.; Gutiérrez, A.; Palos, R.; Azkoiti, M. J.; Arandes, J. M.; Bilbao, J. Cracking of Scrap Tires Pyrolysis Oil in a Fluidized Bed Reactor under Catalytic Cracking Unit Conditions. Effects of Operating Conditions. Energy Fuels 2019, 33, 3133–3143.

3. Hanping, C.; Bin, L.; Haiping, Y.; Guolai, Y.; Shihong, Z. Experimental Investigation of Biomass Gasification in a Fluidized Bed Reactor. Energy Fuels 2008, 22, 3493–3498.

4. Fan, X.; Zhou, C. Estimation of Bed Expansion and Separation Density of Gas–Solid Separation Fluidized Beds Using a Micron-Sized-Particle-Dense Medium. Separations 2021, 8, No. 242.

5. Basu, P. Combustion of Coal in Circulating Fluidized-Bed Boilers: A Review. Chem. Eng. Sci. 1999, 54, 5547–5557.

6. Matsuoka, S.; Maruyama, H.; Matsuuya, T.; Ghadiri, M. Triboelectric Charging of Powders: A Review. Chem. Eng. Sci. 2010, 65, 5781–5807.
(7) Mohrani, P.; Bi, H. T.; Grace, J. R. Electrostatic Charge Generation in Gas-Solid Fluidized Beds. J. Electrostat. 2005, 63, 165–173.

(8) Sowinski, A.; Mayne, A.; Mehrani, P. Effect of Fluidizing Particle Size on Electrostatic Charge Generation and Reactor Wall Fouling in Gas-Solid Fluidized Beds. Chem. Eng. Sci. 2012, 71, 552–563.

(9) Lu, J.; Fujii, S.; Yasuda, M.; Matsuoka, S. Analysis of Wall Fouling and Electrostatic Charging in Gas-Solid Fluidized Beds. Adv. Powder Technol. 2020, 31, 3485–3491.

(10) Hendrickson, G. Electrostatics and Gas Phase Fluidized Bed Polymerization Reactor Wall Sheeting. Chem. Eng. Sci. 2006, 61, 1041–1064.

(11) Park, A.-H. A. Electrostatic Charging in Gas—Solid Fluidized Beds; University of British Columbia, 2000.

(12) Fotovat, F.; Bi, X. T.; Grace, J. R. A Perspective on Electrostatics in Gas-Solid Fluidized Beds: Challenges and Future Research Needs. Powder Technol. 2018, 329, 65–75.

(13) Guardiola, J.; Rojo, V.; Ramos, G. Influence of Particle Size, Fluidization Velocity and Relative Humidity on Fluidized Bed Electrostatics. J. Electrostat. 1996, 37, 1–20.

(14) Boland, D.; Geldart, D. Electrostatic Charging in Gas Fluidised Beds. Powder Technol. 1972, 5, 289–297.

(15) Saleh, K.; Traore Ndama, A.; Guignon, P. Relevant Parameters Involved in Tribocharging of Powders during Dilute Phase Pneumatic Transport. Chem. Eng. Res. Des. 2011, 89, 2582–2597.

(16) McCarty, L. S.; Whitesides, G. M. Electrostatic Charging Due to Separation of Ions at Interfaces: Contact Electrostriction of Ionic Electrets. Angew. Chem., Int. Ed. 2008, 47, 2188–2207.

(17) Wu, J.; Bi, H. T. Addition of Fines for the Reduction of Powder Charging in Particle Mixers. Adv. Powder Technol. 2011, 22, 332–335.

(18) Park, A. A.; Bi, H. T.; Grace, J. R.; Chen, A. Modeling Charge Transfer and Induction in Gas-Solid Fluidized Beds. J. Electrostat. 2002, 55, 135–158.

(19) He, C.; Bi, X. T.; Grace, J. R. Monitoring Electrostatics and Hydrodynamics in Gas-Solids Bubbling Fluidized Beds Using Novel Electrostatic Probes. Ind. Eng. Chem. Res. 2015, 54, 8333–8343.

(20) Moughrabiah, W. O.; Grace, J. R.; Bi, X. T. Effects of Pressure, Temperature, and Gas Velocity on Electrostatics in Gas-Solid Fluidized Beds. Ind. Eng. Chem. Res. 2009, 48, 320–325.

(21) Park, A.-H.; Bi, H.; Grace, J. R. Reduction of Electrostatic Charges in Gas—Solid Fluidized Beds. Chem. Eng. Sci. 2002, 57, 153–162.

(22) Chew, J. W.; Cocco, R. A. Application of Machine Learning Methods to Understand and Predict Circulating Fluidized Bed Riser Flow Characteristics. Chem. Eng. Sci. 2020, 217, No. 115503.

(23) Patel, A. M.; Cocco, R. A.; Chew, J. W. Key Influence of Clusters of Geldart Group B Particles in a Circulating Fluidized Bed Riser. Chem. Eng. J. 2021, 413, No. 127386.

(24) Zhou, J.; Liu, D.; Ye, M.; Liu, Z. Data-Driven Prediction of Minimum Fluidization Velocity in Gas-Fluidized Beds Using Data Extracted by Text Mining. Ind. Eng. Chem. Res. 2021, 60, 13727–13739.

(25) Guo, Q.; Ye, M.; Yang, W.; Liu, Z. A Machine Learning Approach for Electrical Capacitance Tomography Measurement of Gas—Solid Fluidized Beds. AIChE J. 2019, 65, No. e16583.

(26) Grisaifi, A.; Fabrizio, A.; Meyer, B.; Wilkins, D. M.; Corinobouef, C.; Ceriotti, M. Transferable Machine-Learning Model of the Electron Density. ACS Cent. Sci. 2019, 5, 57–64.

(27) Rostami, A.; Anbaz, M. A.; Erfani Gahrooei, H. R.; Arabloo, M.; Bahadori, A. Accurate Estimation of CO2 Adsorption on Activated Carbon with Multi-Layer Feed-Forward Neural Network (MLFFNN) Algorithm. Egypt. J. Pet. 2018, 27, 65–73.

(28) Rostami, A.; Kalantari-Meybodi, M.; Karimi, M.; Tatar, A.; Mohammadi, A. H. Efficient Estimation of Hydrolyzed Polyelectrolyte (HPAM) Solution Viscosity for Enhanced Oil Recovery Process by Polymer Flooding. Oil Gas Sci. Technol. 2018, 73, No. 22.

(29) Rostami, A.; Baghban, A.; Shirazian, S. On the Evaluation of Density of Ionic Liquids: Towards a Comparative Study. Chem. Eng. Res. Des. 2019, 147, 648–663.
(54) Smola, A. J.; Schölkopf, B. A Tutorial on Support Vector Regression. Stat. Comput. 2004, 14, 199−222.
(55) Han, S.; Qubo, C.; Meng, H. In Parameter Selection in SVM with RBF Kernel Function, World Automation Congress, 2012; pp 1−4.
(56) Mielniczuk, J.; Tyrecha, J. Consistency of Multilayer Perceptron Regression Estimators. Neural Networks 1993, 6, 1019−1022.
(57) Murtagh, F. Multilayer Perceptrons for Classification and Regression. Neurocomputing 1991, 2, 183−197.
(58) Thomas, P.; Suhner, M. C. A New Multilayer Perceptron Pruning Algorithm for Classification and Regression Applications. Neural Process. Lett. 2015, 42, 437−458.
(59) Anastassiou, G. A. Multivariate Hyperbolic Tangent Neural Network Approximation. Comput. Math. Appl. 2011, 61, 809−821.
(60) Liu, D. C.; Nocedal, J. On the Limited Memory BFGS Method for Large Scale Optimization. Math. Program. 1989, 45, 503−528.
(61) Budescu, D. V. Dominance Analysis: A New Approach to the Problem of Relative Importance of Predictors in Multiple Regression. Psychol. Bull. 1993, 114, 542−551.
(62) Azen, R.; Budescu, D. V. Comparing Predictors in Multivariate Regression Models: An Extension of Dominance Analysis. J. Educ. Behav. Stat. 2016, 31, 157−180.
(63) Farrar, D. E.; Glauber, R. R. Multicollinearity in Regression Analysis: The Problem Revisited. Rev. Econ. Stat. 1967, 49, 92.
(64) O’Brien, R. M. A Caution Regarding Rules of Thumb for Variance Inflation Factors. Qual. Quant. 2007, 41, 673−690.
(65) Sobolı́, I. Global Sensitivity Indices for Nonlinear Mathematical Models and Their Monte Carlo Estimates. Math. Comput. Simul. 2001, 55, 271−280.
(66) Saltelli, A. Making Best Use of Model Evaluations to Compute Sensitivity Indices. Comput. Phys. Commun. 2002, 145, 280−297.
(67) Saltelli, A.; Annoni, P.; Azzini, I.; Campolongo, F.; Ratto, M.; Tarantola, S. Variance Based Sensitivity Analysis of Model Output. Design and Estimator for the Total Sensitivity Index. Comput. Phys. Commun. 2010, 181, 259−270.