Supporting Information

Predicting Bulk Density for Agglomerated Raspberry Ketone via Integrating Morphological and Size Metrics using Artificial Neural Networks

Xiaomeng Zhou [1], Shutian Xuanyuan [1], Yang Ye [1], Ying Sun [1], Haowen Du [1], Luguang Qi [1], Chang Li [1] and Chuang Xie [2,*]

[1] School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China; 2021207594@tju.edu.cn (X.Z.); xuany@tju.edu.cn (S.X.); yeyang96@tju.edu.cn (Y.Y.); sy7998@tju.edu.cn (Y.S.); 2021207572@tju.edu.cn (H.D.); qiluguang@tju.edu.cn (L.Q.); chang_li@tju.edu.cn (C.L.)

[2] School of Chemical Engineering and Technology, National Engineering Research Center of Industrial Crystallization Technology, Tianjin University, Tianjin 300072, China

Table S1. Detailed Experimental Conditions for Seeding Crystallization.

Exp.	Seeding time[a]	wt% Raspberry Ketone[b]	Seed size (μm)	Seeds loading (wt %)
1	Before LLP separation, (the antisolvent was added continuously for 78 minutes)	8.35	75-100	2
2	Before LLP separation, (the antisolvent was added continuously for 78 minutes)	8.35	150-180	2
3	After LLP separation, (the antisolvent was added continuously for 87 minutes)	7.69	75-100	2
4	After LLP separation, (the antisolvent was added continuously for 87 minutes)	7.69	150-180	2

[a] Note: LLPS occurred at 85 min of continuous addition of antisolvent, i.e., the mass fraction of raspberry ketone was 7.81%. LLPS was observed by stopping the stirring and then standing for a period of time (Figure S1). LLPS: liquid-liquid phase separation.
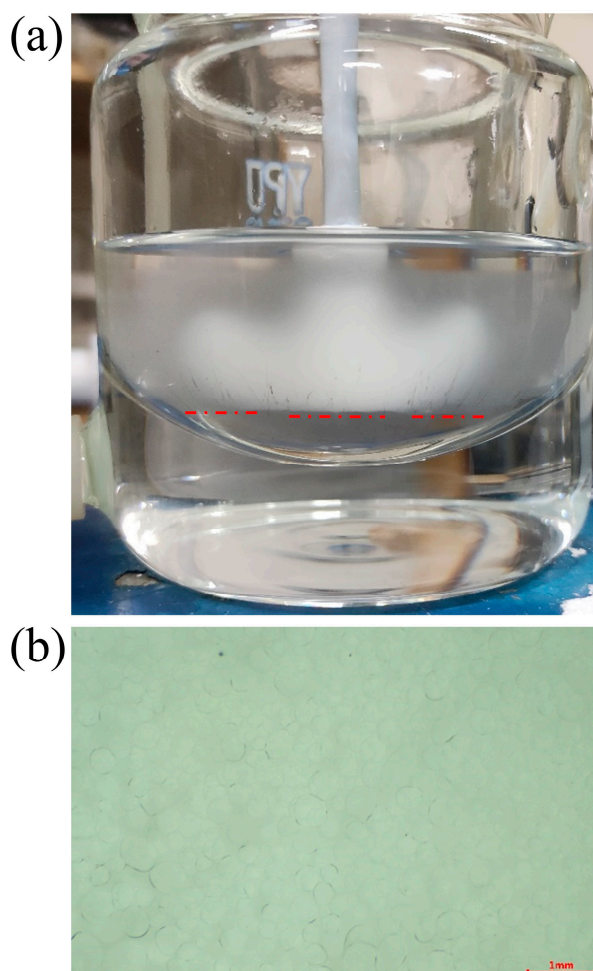
(a)

(b)

Figure S1. (a) LLPS was observed by stopping the stirring and then standing for a period of time. (b) Oil droplets were observed under the microscope.
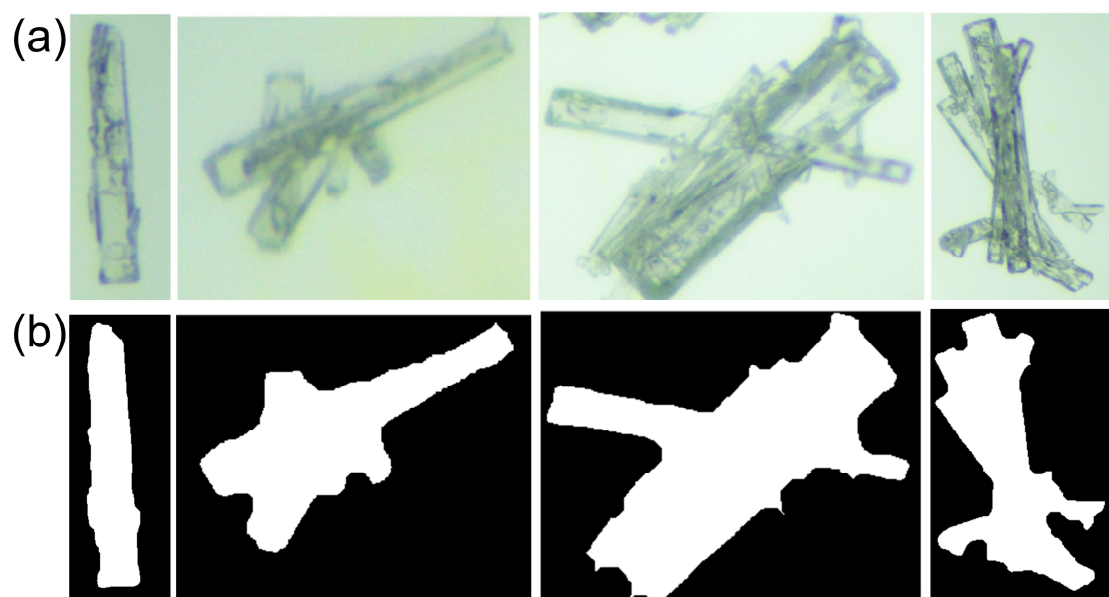
Figure S2. (a) Microscope images of crystals with different degrees of agglomeration. (b)

The images after image processing.

Table S2. Ratios of different particle size fractions based on simplex-lattice design.

Fine	Medium	Coarse
1	0	0
0	1	0
0	0	1
0	1/2	1/2
1/2	0	1/2
1/2	1/2	0
1/6	1/6	2/3
1/6	2/3	1/6
2/3	1/6	1/6
1/3	1/3	1/3

Table S3. The 18 image descriptors extracted and their definitions.

Type	Symbol	Descriptor	Uniit	Description
Size (Dimensioned image descriptors describing the particle's size)	A	Area	mm^2	Projection area of the particles
	P	Perimeter	mm	Boundary length of the particles
	D_{eq}	Equivalent diameter	mm	Diameter of a coextensive circle, $D_{eq}=2\cdot\sqrt{A/\pi}$
	F_{max}	Maximal Feret diameter	mm	Maximal Feret diameter
	F_{min}	Minimal Feret diameter	mm	Minimal Feret diameter
Shape (Dimensionless image descriptors describing the particle's shape)	c	Circularity	-	Similarity with a circle, $c=4\cdot\pi\cdot A/P^2$
	r	Roundness	-	Squared ratio of Equivalent diameter and major ellipse axis, $r=(D_{eq}/L)^2$
	s	Solidity	-	Ratio of the projection area of the particles to the convex hull area, $s=A/A_{convex}$
	e	Elongation	-	Ratio of maximum Ferret diameter to minimum Ferret diameter, $e=F_{max}/F_{min}$
	α_r	Aspect ratio	-	Ratio of the long axis to the short axis of the best-fit ellipse, $\alpha_r=L/B$
	α_{LB}	Convex box ratio	-	Ratio between convex hull area and ellipse axes, $\alpha_{LB}=A_{convex}/L\cdot B$
Roughness (Dimensionless image descriptors describing the roughness of the contour)	I_{CAV}	Concavity index	-	Ratio of the maximum concave area of the projected particle to the projection area, $I_{CAV}=A_{concave}/A$
		Concavity	-	Ratio of the particle projection area to the concave area, $=A/A_{concave}$
		Convexity	-	Ratio of the convex hull perimeter of the particles to the boundary length, $=P_{convex}/P$
	CDR	Concave defect ratio	-	Ratio of the maximum concavity depth of the particle to the maximum Ferret diameter, $CDR=d_{max}\cdot D_{eq}/F_{max}$
	n_{coc}	Num. conc.	-	Number of concavity points
	d_{max}	Max depth	-	Scaled max. concavity depth
	N	Number of primary particles	-	Number of primary particles embedded in agglomerates

Table S4. Algorithm for Auto-tuning of parameters in ANNs.

Step 1	Initialize the number of hidden layers, the number of nodes, and the activation function (1 represents tansigmoid, 2 represents logsigmoid)[a].
Step 2	Set the population count and generation count and start NSGA-II.
Step 3	For a given population, determine the network architecture and the activation function.
Step 4	Initialize the neural network weights and biases, which are then assigned to the created neural network.
Step 5	Train and validate the ANN using Back Propagation and LMA.
Step 6	Test the ANN using the test data and evaluate R2.
Step 7	Determine the number of parameters.
Step 8	Repeat Steps 3 to 7 for all individuals in the current generation.
Step 9	Update the population by performing crossover, mutation and selection.
Step 10	Run all generations or until NSGA-II converges.

[a] Commonly used nonlinear activation functions：

$$\text{logsigmoid:} \quad f(x) = \frac{1}{1 + exp^{(-x)}} \tag{S1}$$

$$\text{tansigmoid:} \quad f(x) = \frac{2}{1+exp^{(-2x)}} - 1 \tag{S2}$$

[b] ANN: Artificial Neural Network. LMA: Levenberg-Marquardt algorithm.

Table S5. The image descriptors represented by numbers 1 to 18.

Number	Image descriptor
1	Num. conc.
2	Max depth
3	Number of primary particles
4	Solidity
5	Convexity
6	Concavity
7	Concavity index
8	Concave defect ratio
9	Elongation
10	Convex box ratio

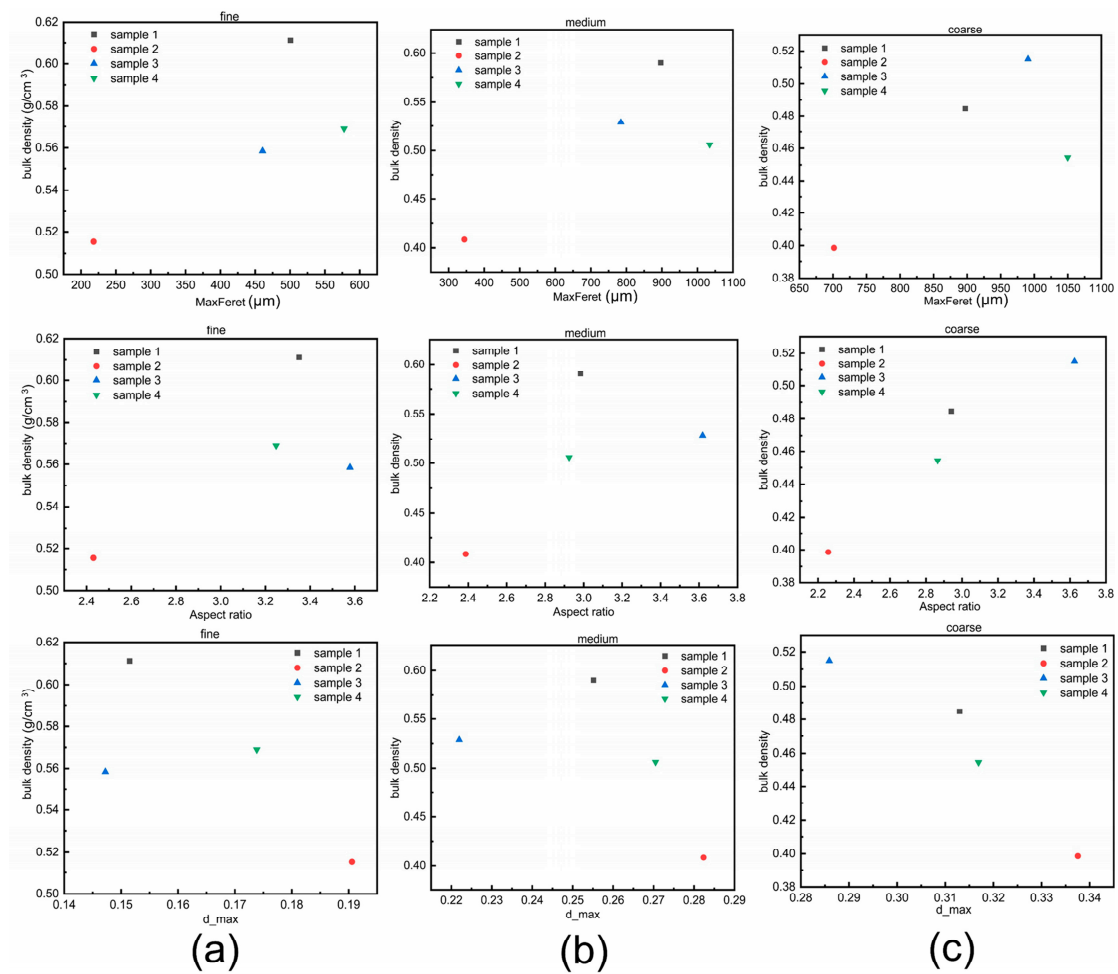| 11	Aspect ratio
12	Circularity
13	Roundness
14	Area
15	Equivalent diameter
16	Maximal Feret diameter
17	Minimal Feret diameter
18	Perimeter

Figure S3. The mean values of the image descriptors for fine (a), medium (b), and coarse (c) particles of the four samples versus their tapped bulk densities. Samples 1 and 2 were obtained under the conditions of adding small sized (75-100 μm) and large sized (150-180 μm) seeds, respectively, before LLPS, samples 3 and 4 were obtained under the conditions

of adding small sized (75-100 μm) and large sized (150-180 μm) seeds, respectively, after LLPS.
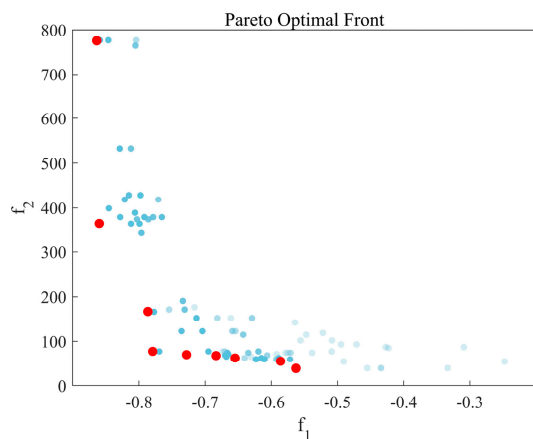

Figure S4. The evolution of pareto front with iterative process. The functions f_1 and f_2 represent the prediction accuracy and the number of model parameters, respectively. The red dots represent the latest Pareto solutions. The pale points indicate earlier solutions, and the darker points are from more recent iterations of the NSGA-II algorithm.

Table S6. ANN model architecture on Pareto front (training with 12 descriptors).

Architecture	Activation Function	Number of Parameters	R2	AIC
[36-7-5-10-1]	2	370	0.8903	956.75
[36-6-5-10-1]	1	328	0.8815	878.01
[36-5-8-3-1]	1	264	0.8773	755.36
[36-5-7-1]	1	235	0.8404	758.01
[36-4-7-3-1]	1	211	0.8303	761.88
[36-5-1]	2	190	0.8214	767.65
[36-4-5-1]	1	179	0.8199	764.12
[36-3-2-8-1]	2	152	0.7953	764.27
[36-3-7-1]	2	147	0.7437	797.16

[36-3-6-1]	2	142	0.7326	797.35
[36-2-13-1]	2	127	0.6682	814.25
[36-2-11-1]	2	119	0.6231	823.06
[36-2-8-1-1]	2	109	0.5594	845.57
[36-1-7-6-1]	1	106	0.5304	847.01

Akaike Information Criterion(AIC):

This work selected an architecture from the Pareto-optimal solutions by AIC, which prevents the over-fitting. The formulation of AIC is shown in Eq. S3 where S is sample size and N is number of parameters in ANN and MSE is mean square error. Among all the solutions, the one with least AIC is selected.
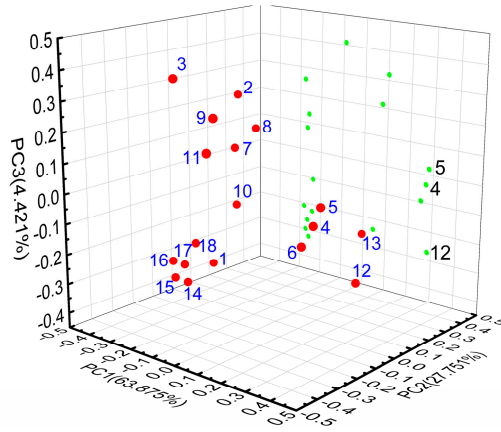
$$AIC = S * log(\text{MSE}) + 2 * N \tag{S3}$$

The detailed method for selecting the descriptors:

In this work, the image descriptors were selected mainly based on the relative size of the loadings of the first three principal components. Figure S5 shows the loading plots of the first three principal components. The specific selection method is as follows:

(1) Firstly, the descriptors with relatively high loadings (1, 2, 4, 5, 12, 14, 17, 18) are selected according to the first principal component (as shown in Figure 6b).

(2) Then the descriptors with relatively high loadings (3, 9, 11, 13, 15, 16) are selected according to the second principal component (as shown in Figure 6b).

(3) Combined with the third principal component (as shown in Figure S5), the descriptors 4 and 5 can be seen to have the relatively lowest loadings (The blue points). In addition, as

can be seen in Figure 1, point 12 has significant variability with points 4, 5, and 6, so point 12 is retained in this work, without considering descriptors 4 and 5.
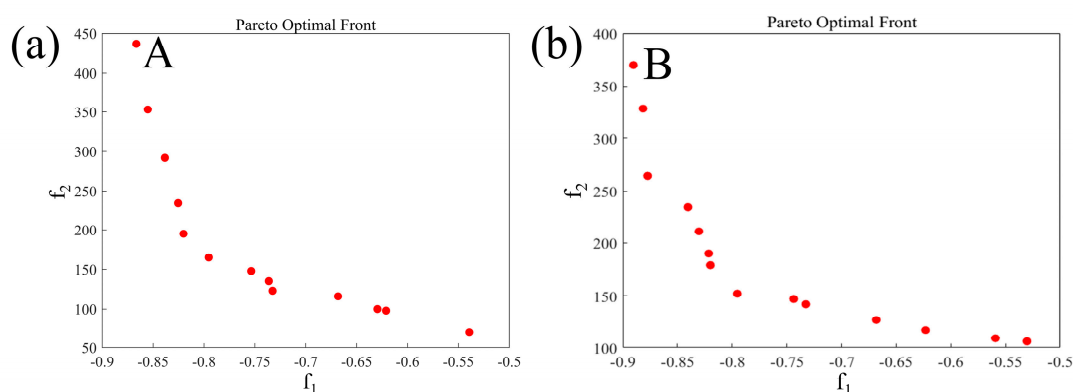

Figure S5. PCA loading plot for the first three principal components. The red points indicate the original data and the blue points indicate the projection on the xz plane.

However, among the 12 descriptors selected, 14 and 18, 3 and 13, and 12 and 17 are somewhat correlated, it is hard to determine whether retaining only {14, 3, 12} is better than {14, 18, 3, 13, 12, 17} from the PCA loading map. Therefore, we compared the effect of using the descriptors {1, 2, 3, 9, 11, 12, 14, 15, 16} and {1, 2, 3, 9, 11, 12, 13, 14, 15, 16, 17, 18} on the accuracy of the prediction model. Figure 3 shows the final Pareto front obtained by training with the descriptors {1, 2, 3, 9, 11, 12, 14, 15, 16} (a) and descriptors {1, 2, 3, 9, 11, 12, 13, 14, 15, 16, 17, 18} (b), respectively.

If prediction accuracy is the single objective, the point A ($R2 = 0.8712$, RMSE = 23.81 kg/m^3) is considered as the best solution in the Pareto front in Figure S6a, whereas the point B ($R2 = 0.8903$, RMSE = 15.02 kg/m^3) is considered as the best solution in Figure S6b, which indicates that keeping the descriptors {18, 13, 17} can improve the accuracy of

the model after training. Therefore, the 12 descriptors {1, 2, 3, 9, 11, 12, 13, 14, 15, 16, 17, 18} were finally selected.

Although the final choice of 12 descriptors is still quite a lot compared to the initial number of 18 descriptors, the highest accuracy of the model was obtained by training.

Figure S6. (a) Final Pareto front obtained by training using the descriptors {1, 2, 3, 9, 11, 12, 14, 15, 16}. (b) Final Pareto front obtained by training using the descriptors {1, 2, 3, 9, 11, 12, 13, 14, 15, 16, 17, 18}. The functions f_1 and f_2 represent the prediction accuracy and the number of model parameters, respectively.