Supplementary Information for

Interpretable machine learning-accelerated seed treatment by nanomaterials for environmental stress alleviation

Hengjie Yu[a,b], Dan Luo[c], Sam Fong Yau Li[d], Maozhen Qu[a,b], Da Liu[a,b], Yingchao He[a,b], and Fang Cheng*[a,b].

[a]College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China

[b]Key Laboratory of Intelligent Equipment and Robotics for Agriculture of Zhejiang Province, Hangzhou 310058, China

[c]Department of Biological and Environmental Engineering, Cornell University, Ithaca, New York 14853, USA

[d]Department of Chemistry, National University of Singapore, 3 Science Drive 3, Singapore 117543, Singapore

* Corresponding author, fcheng@zju.edu.cn (Fang Cheng)

This supplementary information file includes:

- Methods S1 to S5
- Figures S1 to S21
- Tables S1 to S6

Methods

Method S1. Determination of root:shoot ratio

The root:shoot ratio was calculated as the following equation:

$$\text{Root:shoot ratio} = \frac{\text{root dry weight}}{\text{stem dry weight} + \text{leaf dry weight}} \times 100\%$$

Method S2. Determination of SRI

The SRI was calculated as the following equation:

$$SRI = \sum_{i=1}^{10} \frac{k_i}{k_i^c}$$

where k_i is the value of the ith endpoint for each treatment, and k_i^c is the value of the ith endpoint for the control. SRI comprehensively considers the relative value of nano-primed and control treatment in ten biological endpoints, and thus the SRI for control is 10.0.

Method S3. Analytical method, instrument parameters, and data acquisition of metabolomics analysis

The samples were separated by Agilent 1290 infinity LC ultra performance liquid chromatography (UHPLC) on a C-18 column (column temperature:40 ℃). The flow rate of gradient elution was set at 0.4 ml/min, and the injection volume was 2 μL. The mobile phase A = 25 mM ammonium acetate and 0.5% formic acid in water and mobile phase B = methanol. The gradient elution procedure was as follows: 5% B in the first 0.5 min, linearly increased to 100% in the next 9.5 min, and maintained at 100% for 2 min; then it was linearly decreased to 5% in 6 s and maintained at 5% in the next 3.9 min. The sample was placed in an automatic sampler at 4 ℃ during the whole analysis.

Parameter settings of ESI source: Ion Source Gas1 (Gas1) at 60, Ion Source Gas2 (Gas2) at 60, curtain gas at 30, ion source temperature at 600 ℃, ion spray voltage floating at ± 5500 V. The instrument was set to collect data in the m/z range of 60-1000 Da in MS only acquisition, and the accumulation time for TOF MS scan was 0.20 s/spectra. The instrument acquired data over the m/z range 25-1000 Da in auto MS/MS acquisition, and the accumulation time for product ion scan was 0.05 s/spectra. Production scans were acquired using information dependent acquisition with high sensitivity mode selected. The parameters were set as follows: the collision energy at 35 V with ± 15 eV; declustering potential at ± 60 V (positive and negative modes); exclude isotopes within 4 Da and candidate ions to monitor per cycle at 10.

Method S4. Line fitting of SHAP main effects and SHAP interactions

The line fitting was based on python language using "scipy.optimize" and "numpy.polyfit" packages. The fitting equations are described as follows:

(1) Piecewise linear fitting for SHAP main effects of the TEM size of nanoparticles (Fig. 4e):

$$y=\begin{cases}2.00\times10^{-1}x-3.87,\ x\le 21.50\ \text{nm}\\ 1.50\times10^{-3}x+3.90\times10^{-1},\ x>21.50\ \text{nm}\end{cases}$$

where x is the TEM size of nanoparticles, and y is the SHAP main effect value.

(2) Linear fitting for SHAP main effects of the zeta potential of nanoparticles (Fig. 4f):

$$y=-1.14\times10^{-2}x-2.77\times10^{-2}x$$

where x is the zeta potential of nanoparticles, and y is the SHAP main effect value.

(3) Logistic fitting for SHAP interactions between the zeta potential and concentration of nanoparticles (Fig. 4k):

$$y=\begin{cases}\dfrac{-2.28}{1+e^{4.66+5.48\times10^{-1}x_1}}+8.39\times10^{-1},\ x_2=100\ \text{mg/L}\\ \dfrac{2.29}{1+e^{10.57+9.32\times10^{-1}x_1}}-6.91\times10^{-1},\ x_2=50\ \text{mg/L}\end{cases}$$

where x_1 is the zeta potential of nanoparticles, x_2 is the concentration of nanoparticles, and y is the SHAP interaction value.

(4) Plateau fitting for SHAP interactions between the TEM size and concentration of nanoparticles (Fig. 4l):

$$y=\frac{-3.96x}{8.51+x_1}+3.00,\ x_2=100\ \text{mg/L or}\ x_2=50\ \text{mg/L}$$

where x_1 is the TEM size of nanoparticles, x_2 is the concentration of

nanoparticles, and y is SHAP interaction value.

(5) Polynomial fitting for SHAP main effects of the TEM size of nanoparticles (Fig. 5e):

$$y = 1.66\times10^{-6}x^3 - 3.63\times10^{-4}x^2 + 2.44\times10^{-2}x - 4.27\times10^{-1}$$

where x is the TEM size of nanoparticles, and y is the SHAP main effect value.

(6) Polynomial fitting for SHAP main effects of the zeta potential of nanoparticles (Fig. 5f):

$$y = -5.71\times10^{-6}x^3 + 2.49\times10^{-4}x^2 + 1.89\times10^{-3}x - 8.44\times10^{-2}$$

where x is the zeta potential of nanoparticles, and y is the SHAP main effect value.

Method S5. Cost estimate of seed nanopriming (nanoparticles)

Given the following conditions:

ZnO nanoparticle price (30nm, Macklin): RMB ¥ 308 (500 g)

Seed weight: nanoparticle volume: 1:5 g/mL

Planting seeds = 22 kg/ha

Nanoparticle concentration: 200 mg/L

so, we can estimate the nanoparticle fee.

Nanosuspension volume = 22 kg/ha × 5 L/kg = 110 L/ha

Nanoparticle weight = 110 L/ha × 200 mg/L = 22000 mg/ha = 22 g/ha

Nanoparticle fee = 22 g/ha × ¥ 308 ÷ 500 g = ¥ 13.552/ha (around $ 2/ha)

Figures

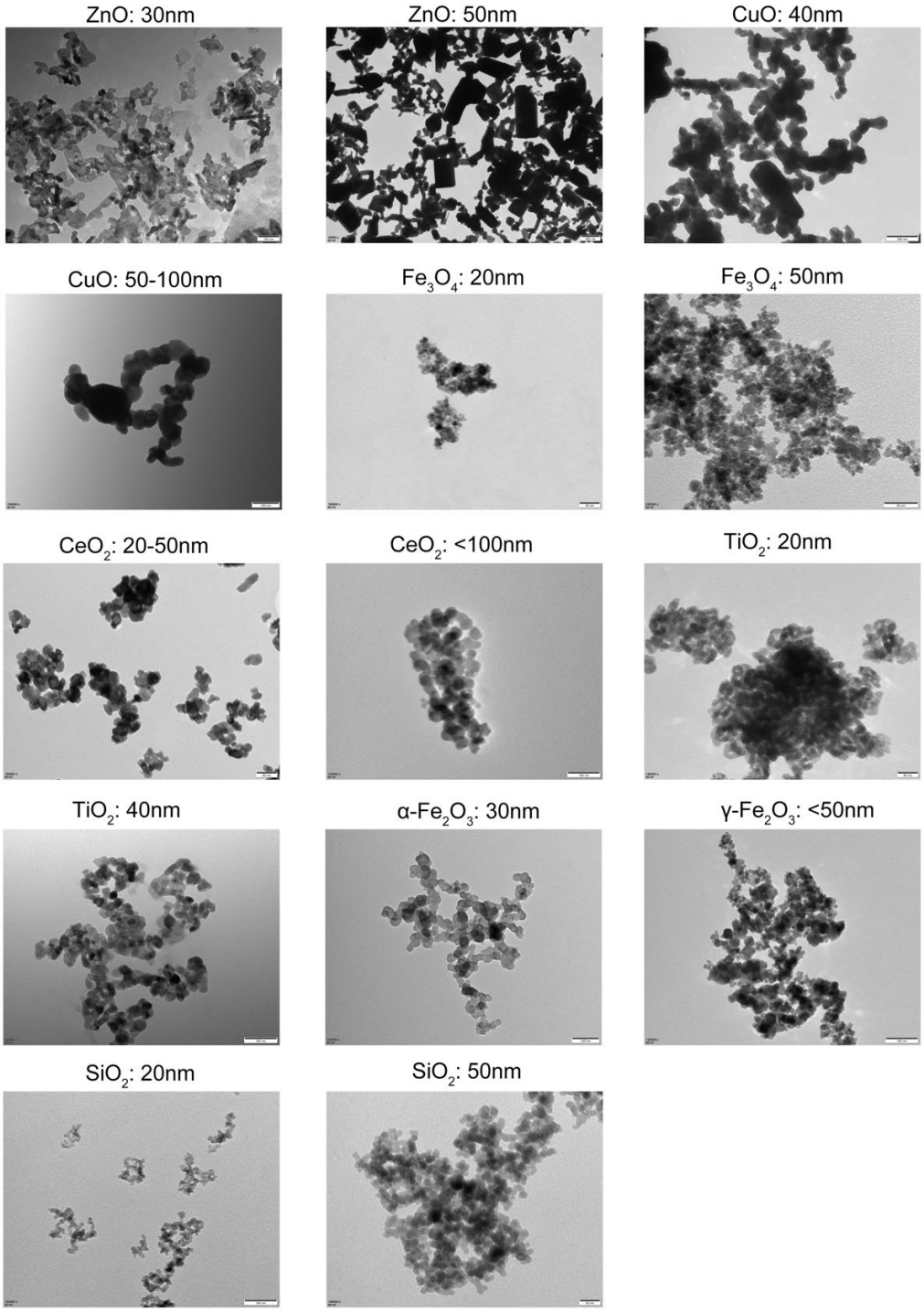

Fig. S1: TEM images of fourteen low-cost metalloid and metal oxide nanoparticles (SiO_2, CeO_2, CuO, Fe_3O_4, ZnO, α-Fe_2O_3, and γ-Fe_2O_3 of different sizes).

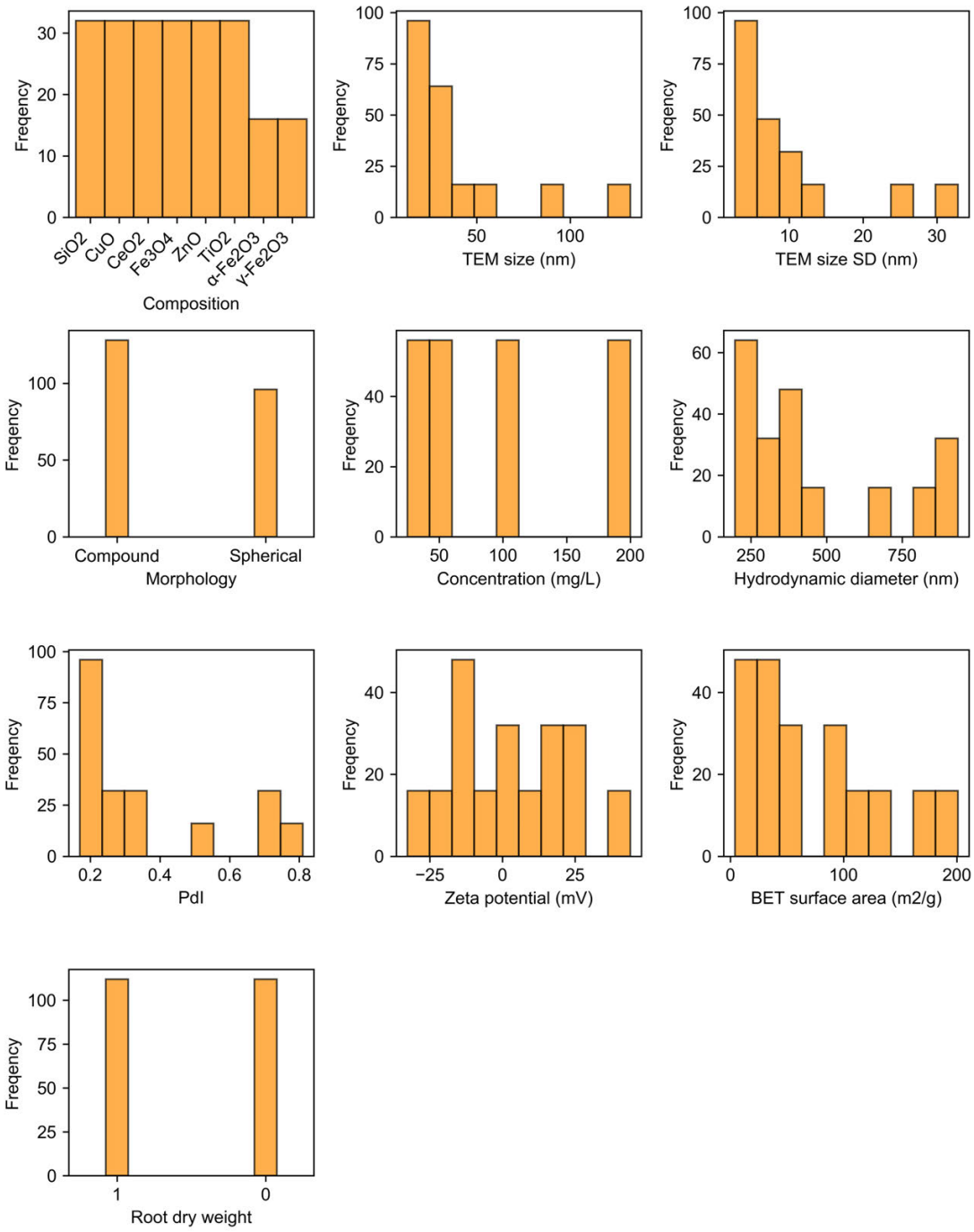

Fig. S2. An overview of the used features and the prediction target (root dry weight).

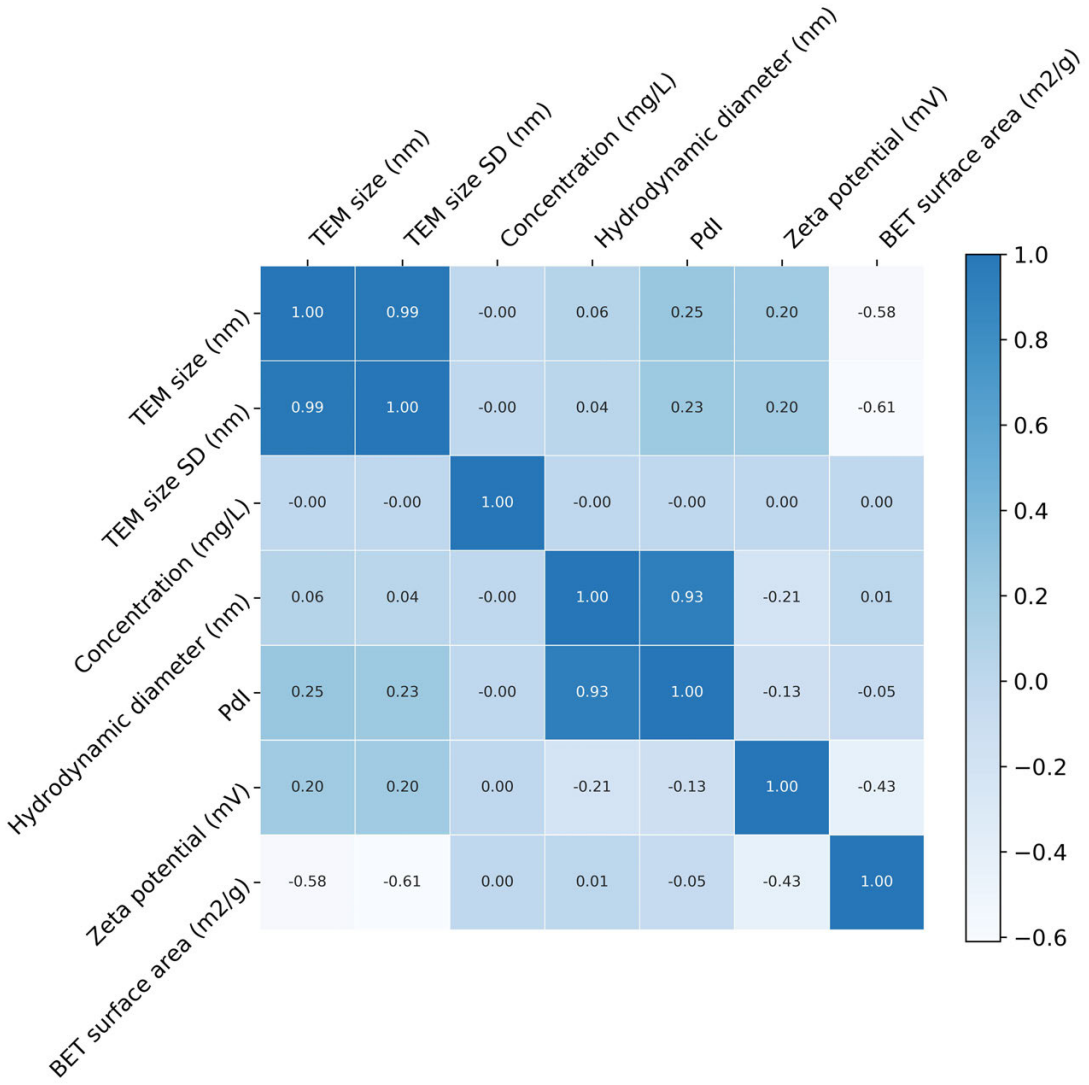

Fig. S3. The heatmap of the Pearson correlation coefficient among numerical factors.

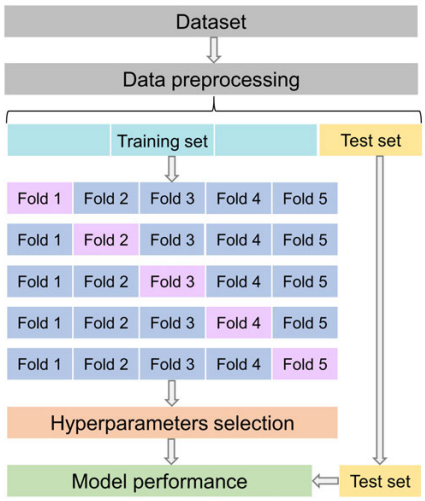

Fig. S4. The workflow for the establishment of the LightGBM models.

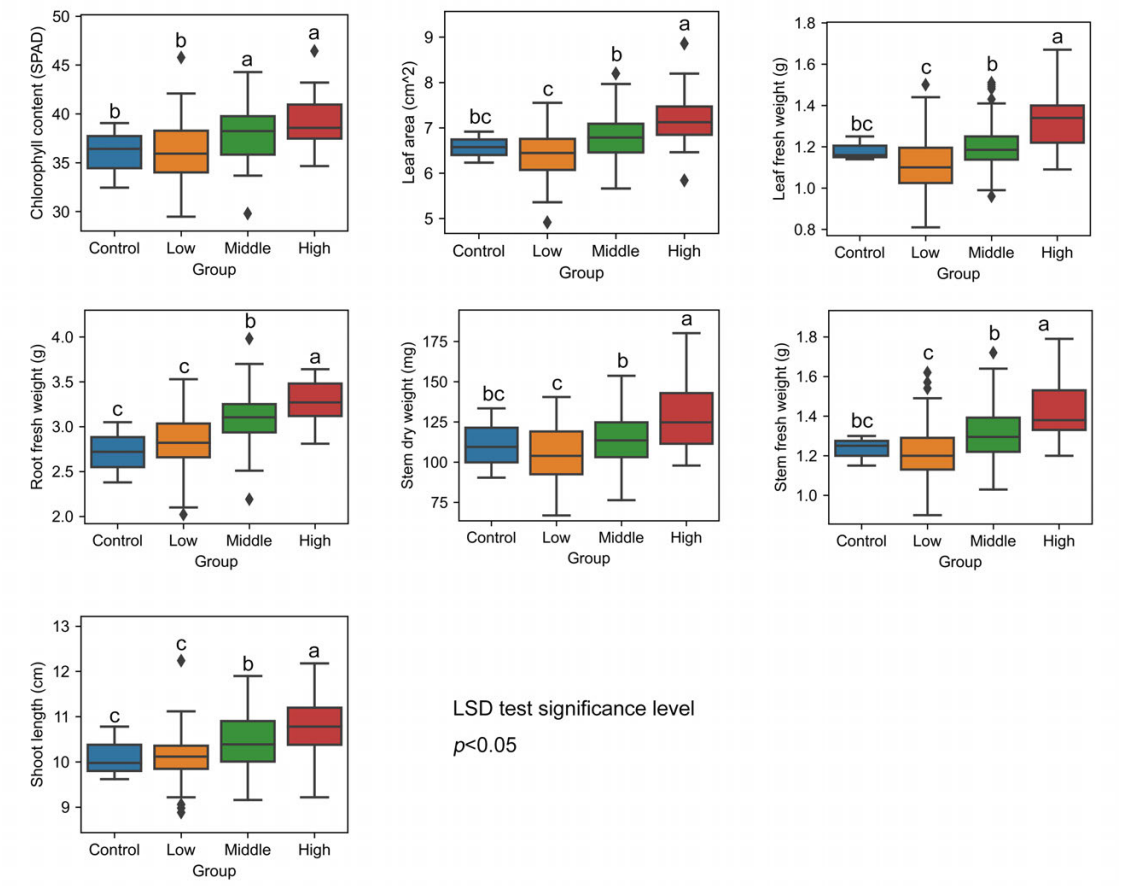

Fig. S5. The differences between three nanopriming groups and the control on biological endpoints under salinity stress.

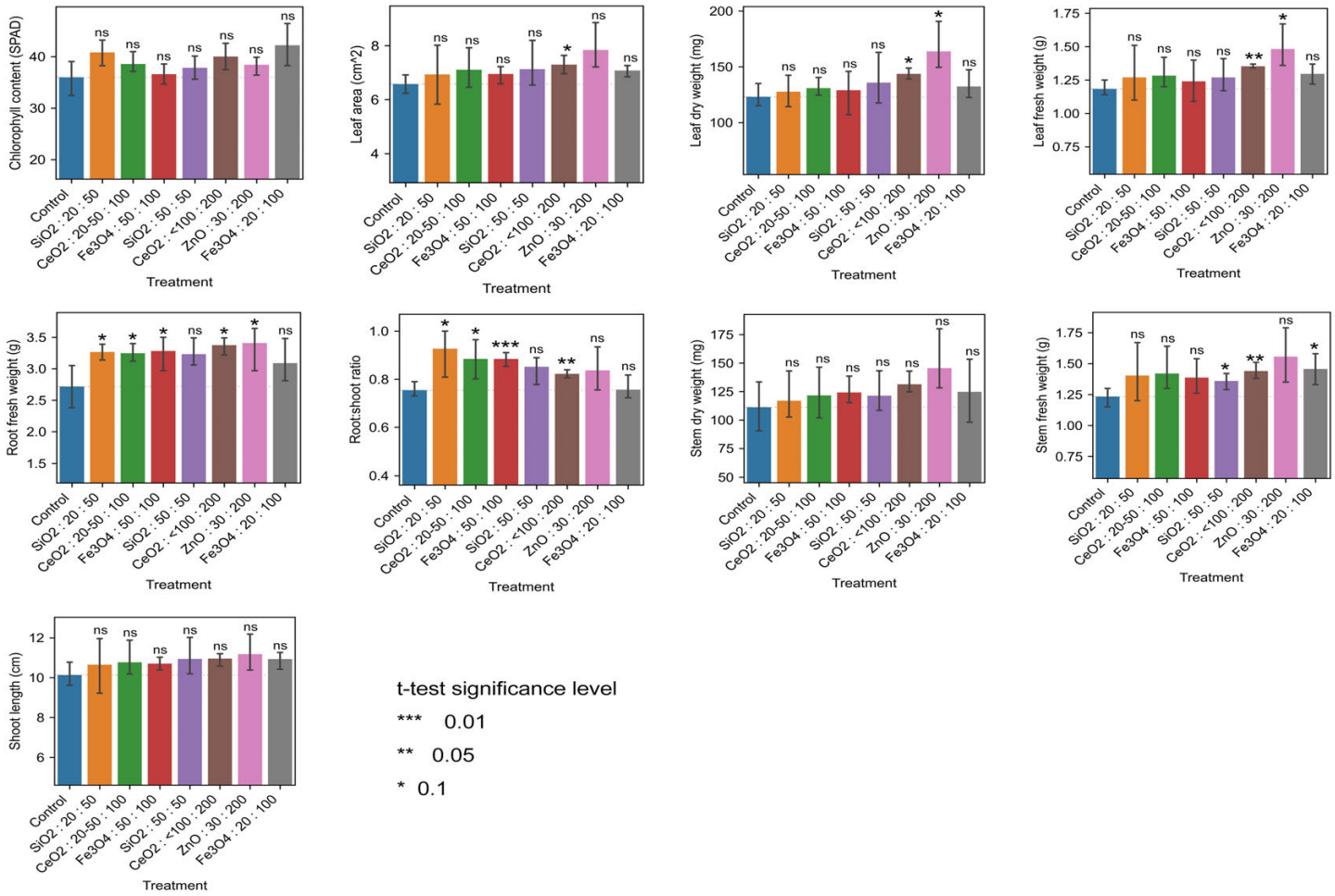

Fig. S6. The comparison of biological endpoints among seven treatments in the High group under salinity stress.

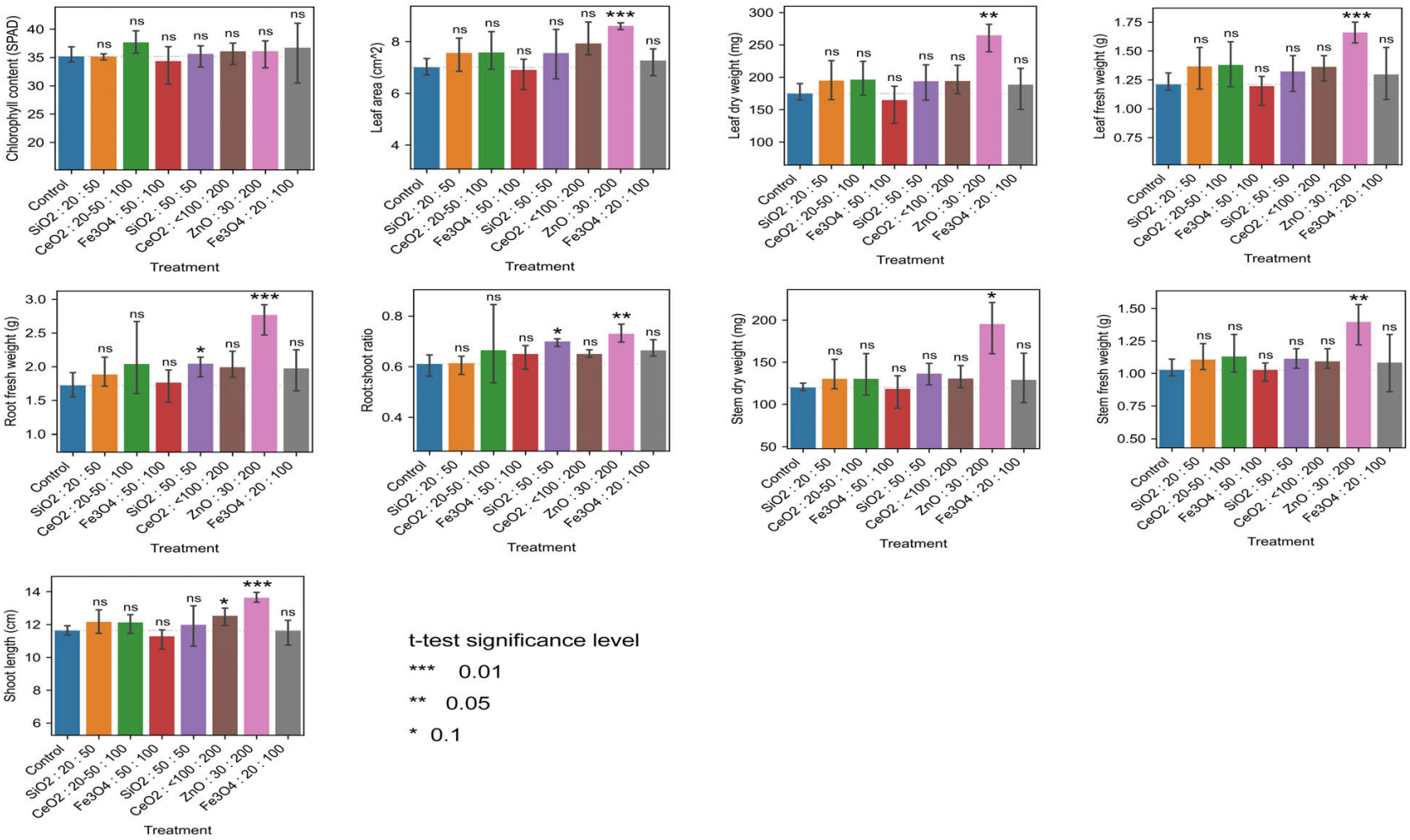

Fig. S7. The comparison of biological endpoints among selected treatments under combined heat-drought stress.

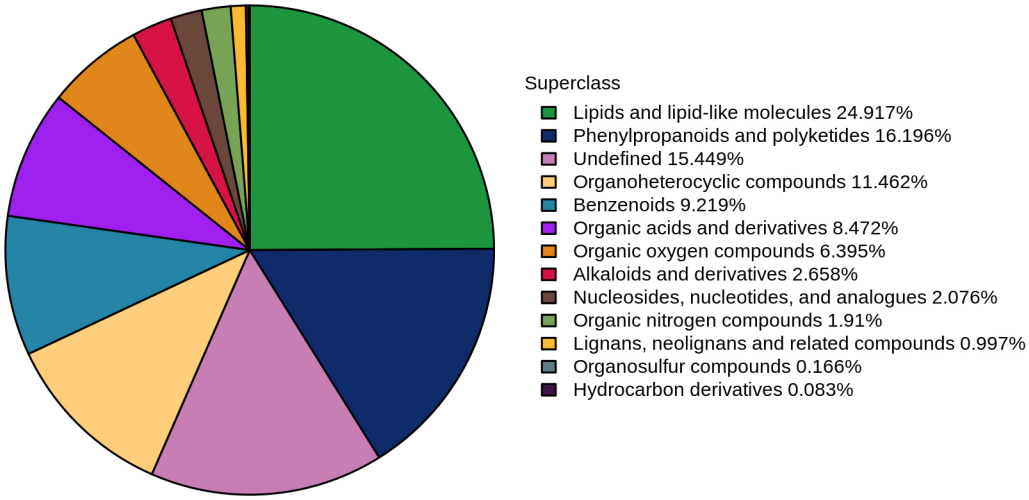

Fig. S8. The superclass of 1204 identified metabolites.

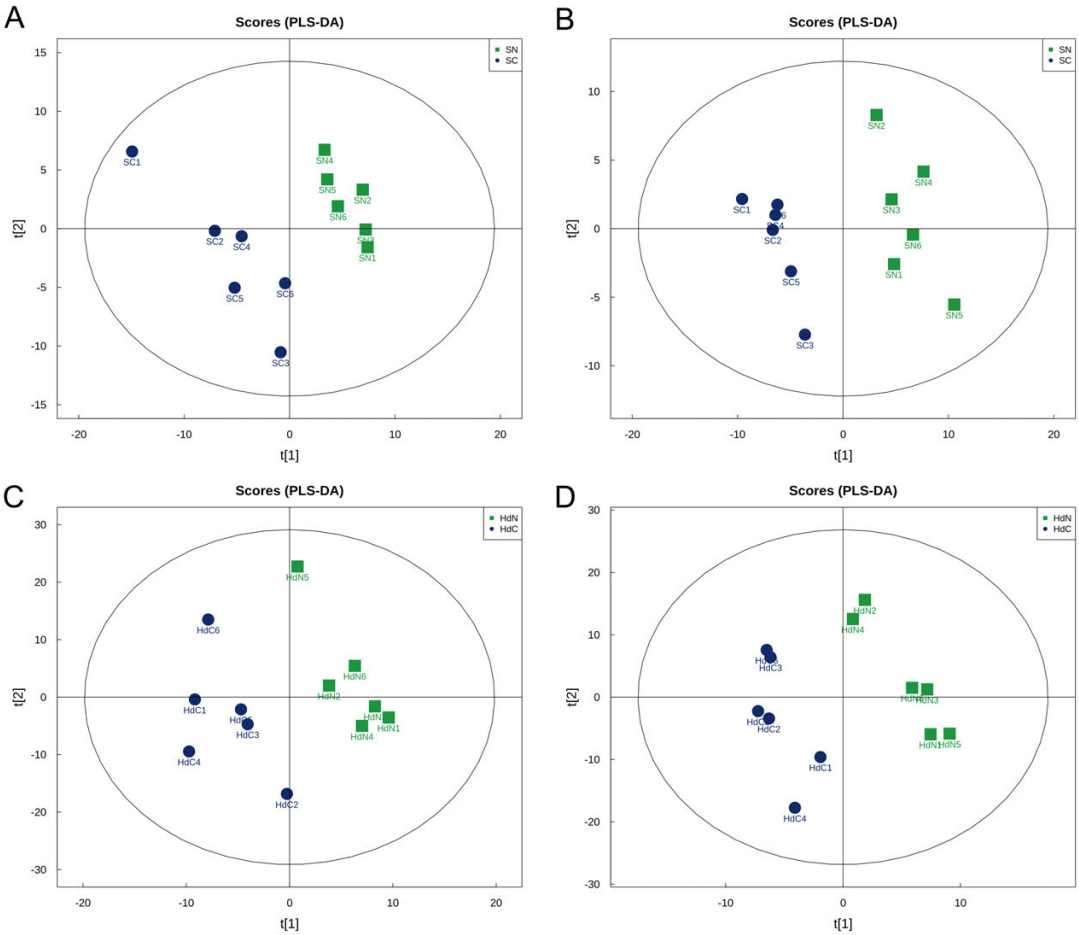

Fig. S9. The score plots of PLS-DA of metabolic profiles in maize leaves after SN and SC seed priming in the positive (A) and negative (B) ion modes. The score plots of PLS-DA of metabolic profiles in maize leaves after HdN and HdC seed priming in the positive (C) and negative (D) ion modes.

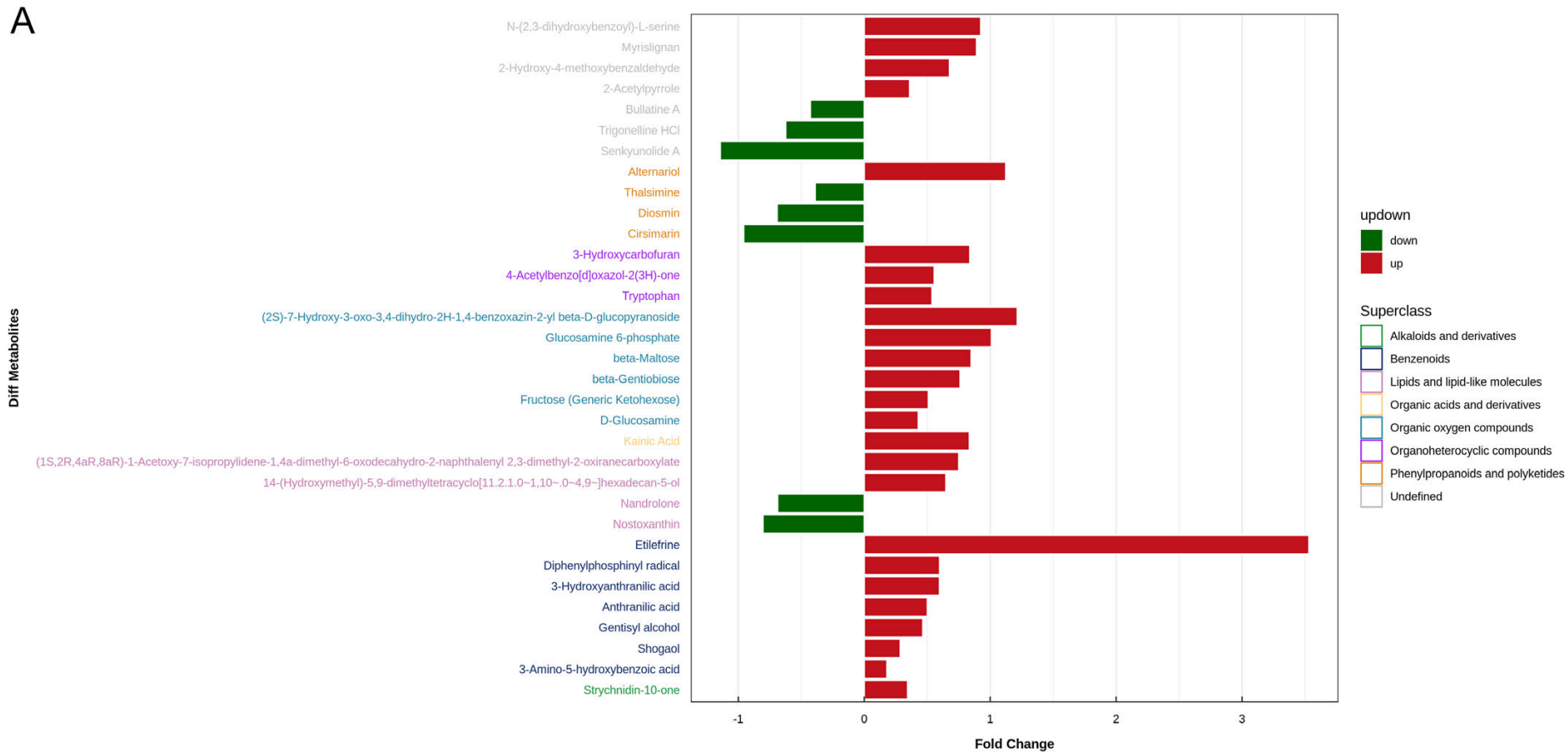

Fig. S10 (A). The up-regulated and down-regulated metabolites in maize leaves after SN and SC seed priming in the positive ion modes.

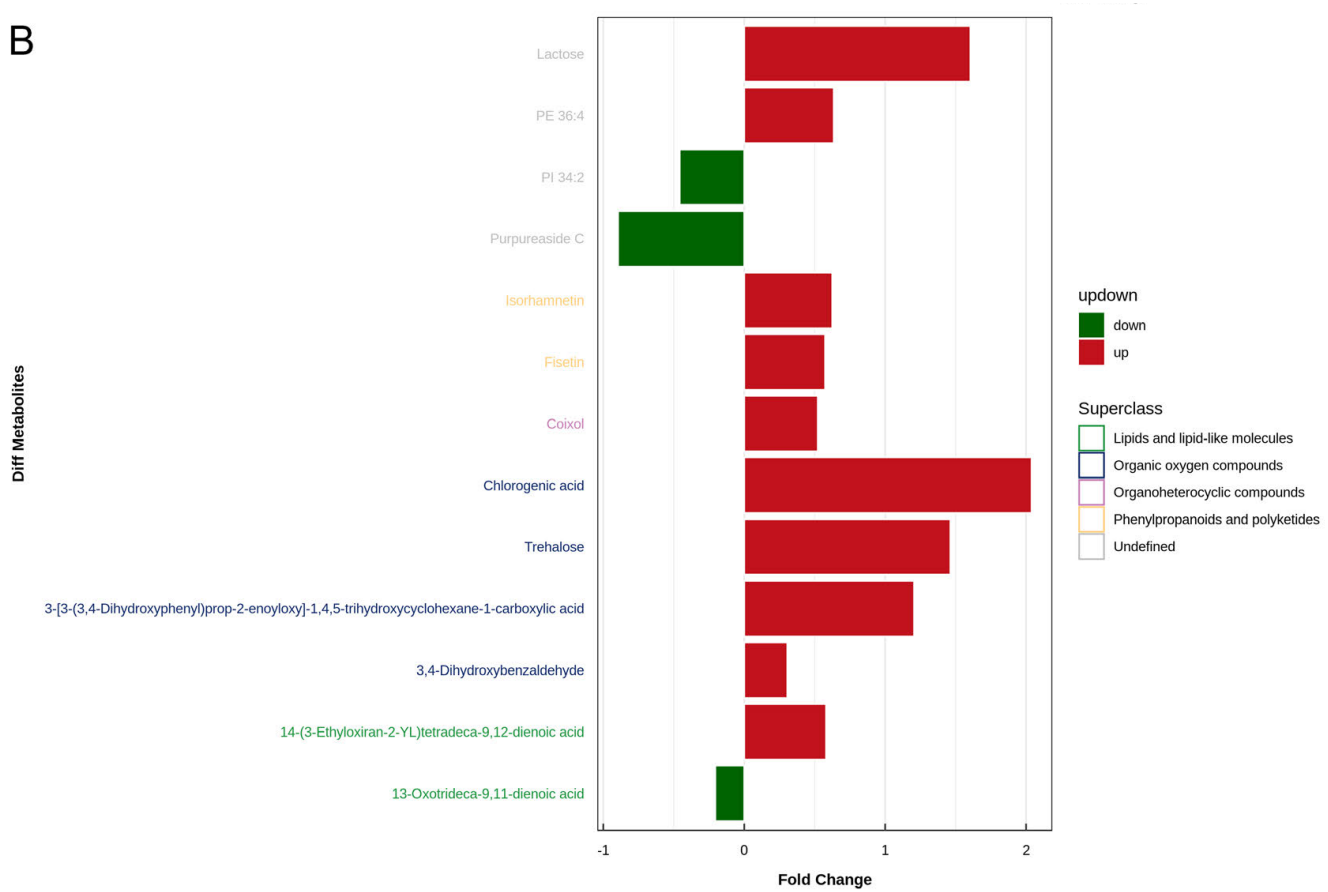

Fig. S10 (B). The up-regulated and down-regulated metabolites in maize leaves after SN and SC seed priming in the negative ion modes.

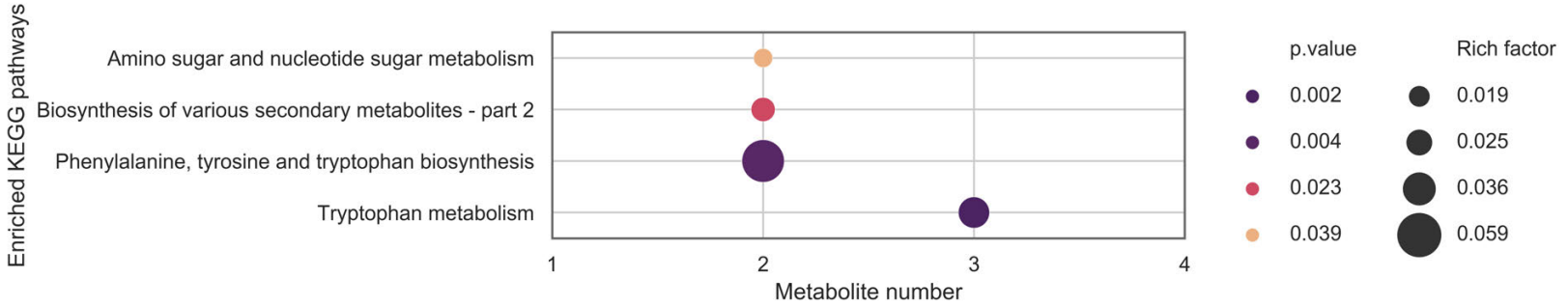

Fig. S11. KEGG pathway enrichment analysis based on significantly different metabolites between SN and SC.

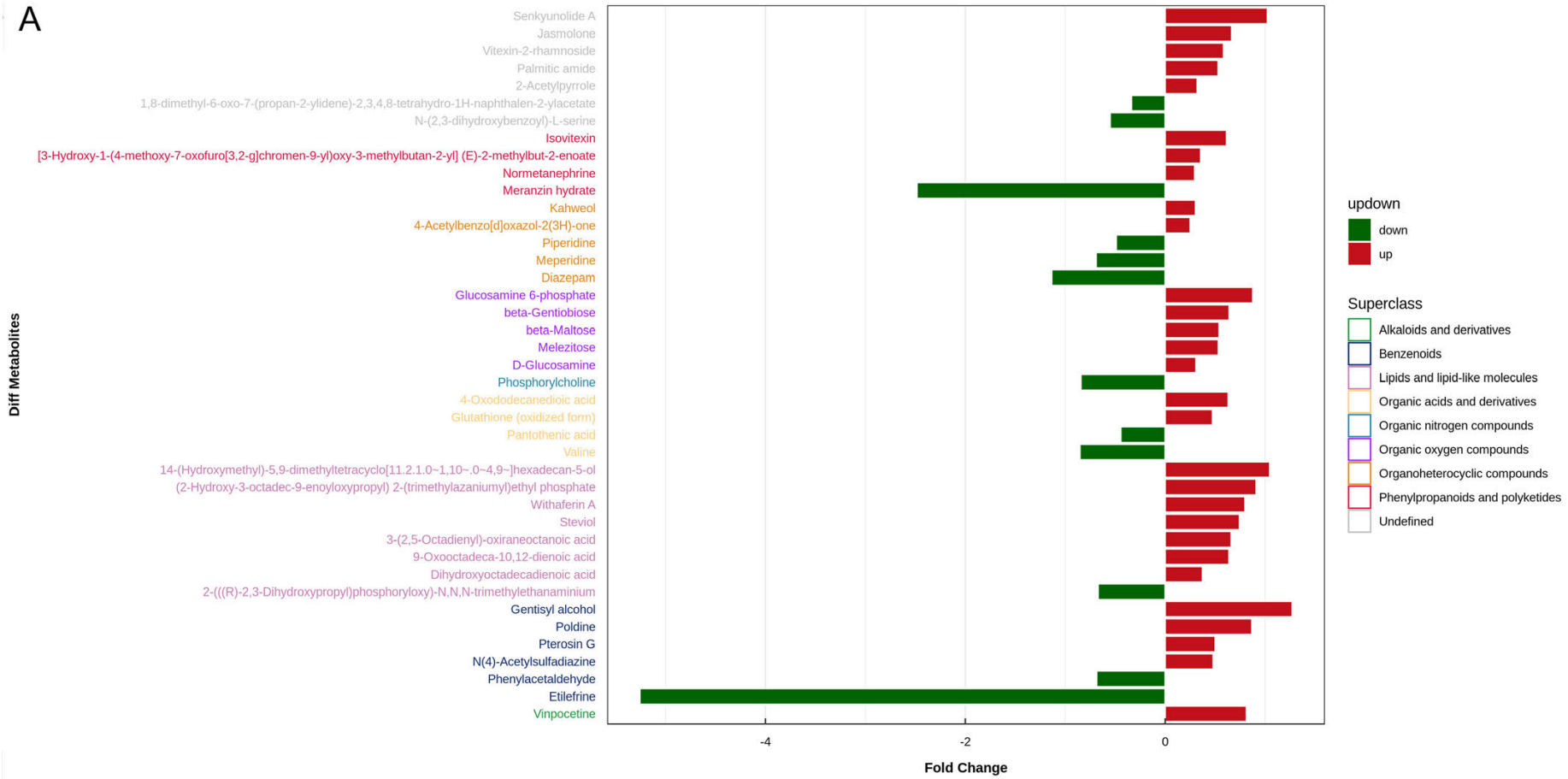

Fig. S12 (A). The up-regulated and down-regulated metabolites in maize leaves after HdN and HdC seed priming in the positive ion modes.

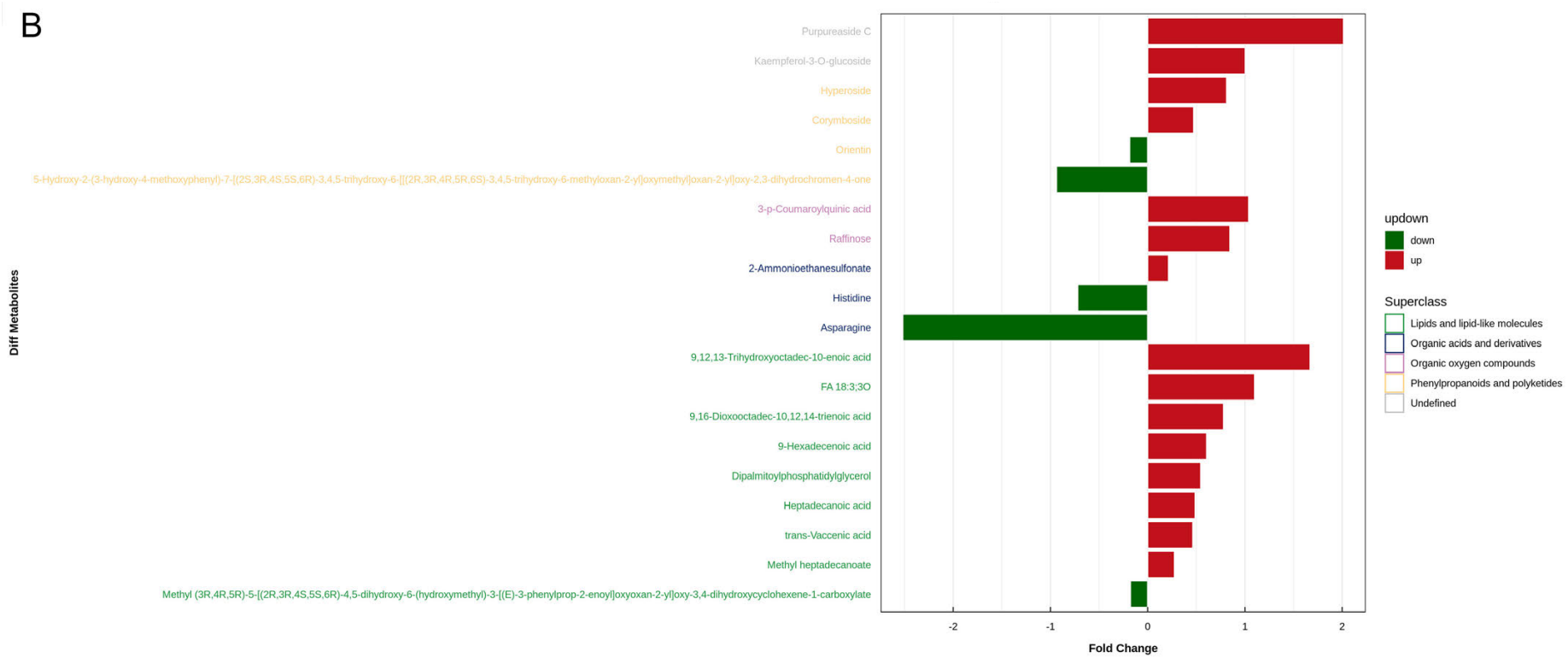

Fig. S12 (B). The up-regulated and down-regulated metabolites in maize leaves after HdN and HdC seed priming in the negative ion modes.

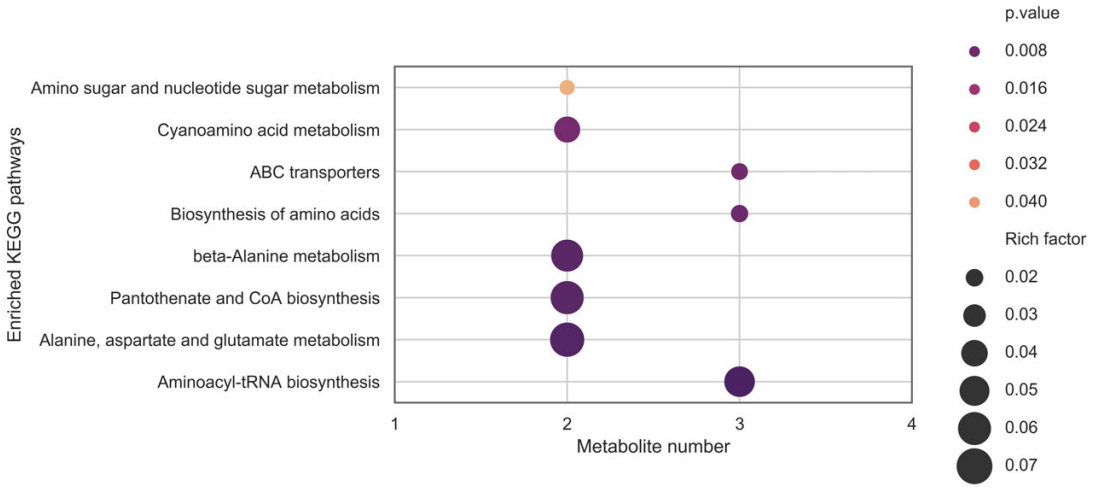

Fig. S13. KEGG pathway enrichment analysis based on significantly different metabolites between HdN and HdC.

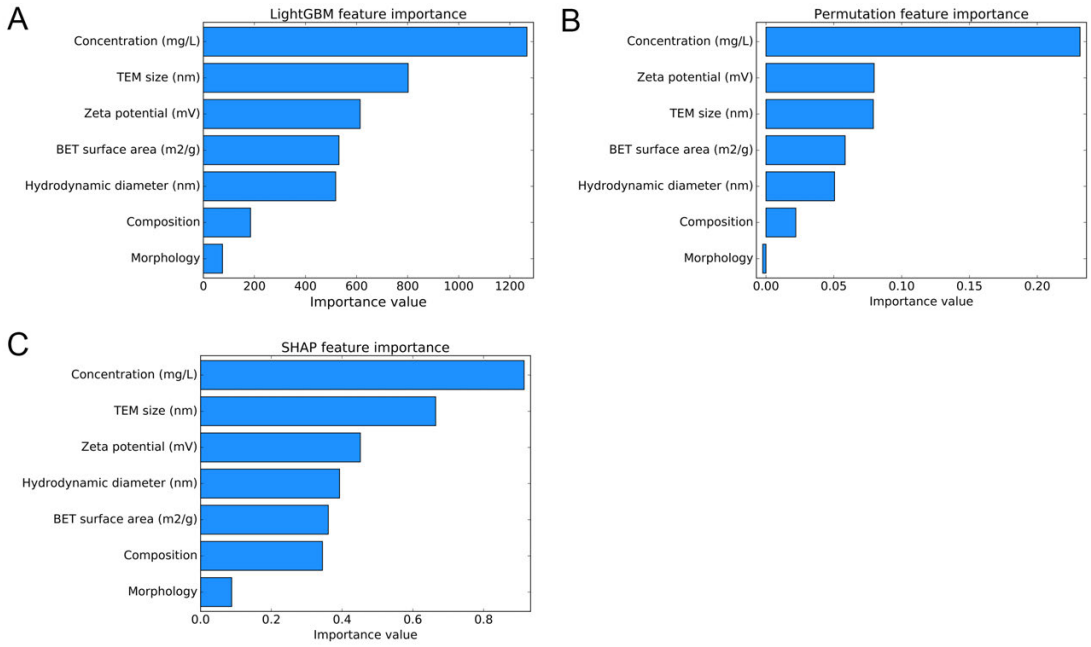

Fig. S14. The absolute values of feature importance obtained by LightGBM feature importance (A), permutation feature importance (B), and SHAP feature importance (C).

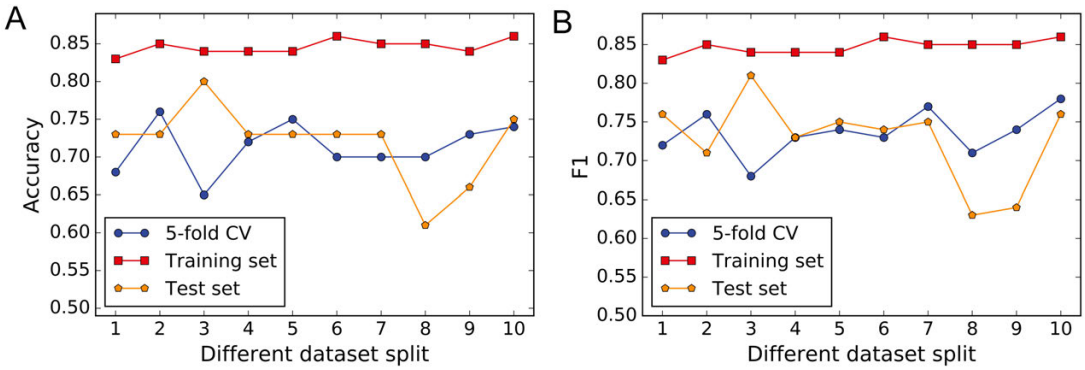

Fig. S15. The accuracy (A) and F1 score (B) of established models on ten dataset splits.

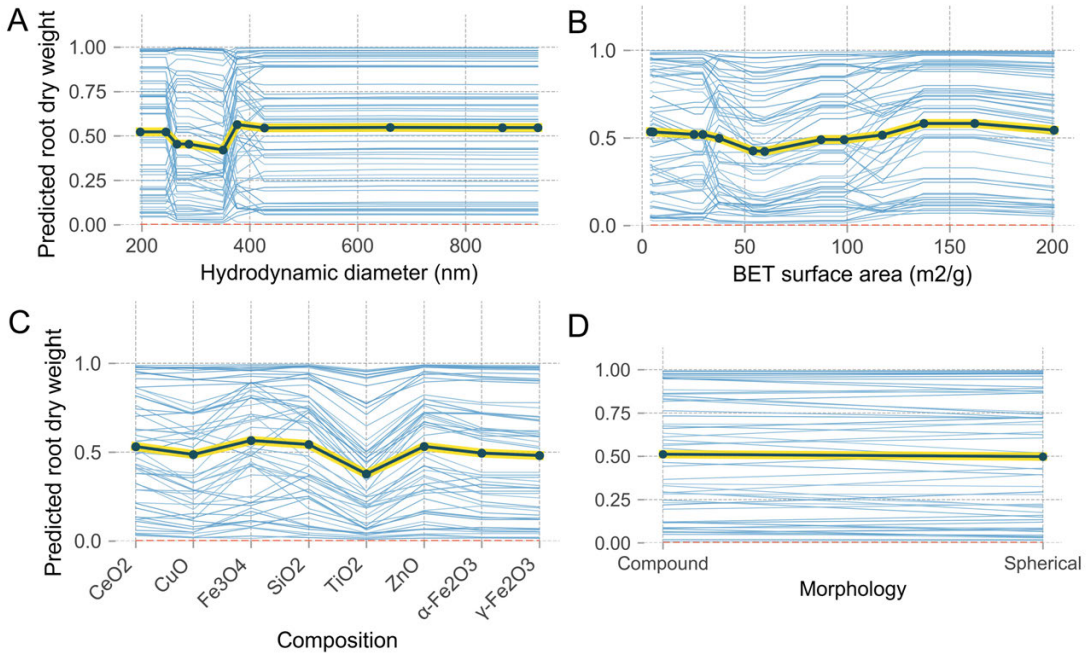

Fig. S16. PDP and ICE plots of hydrodynamic diameter (A), BET surface area (B), composition (C), and morphology (D).

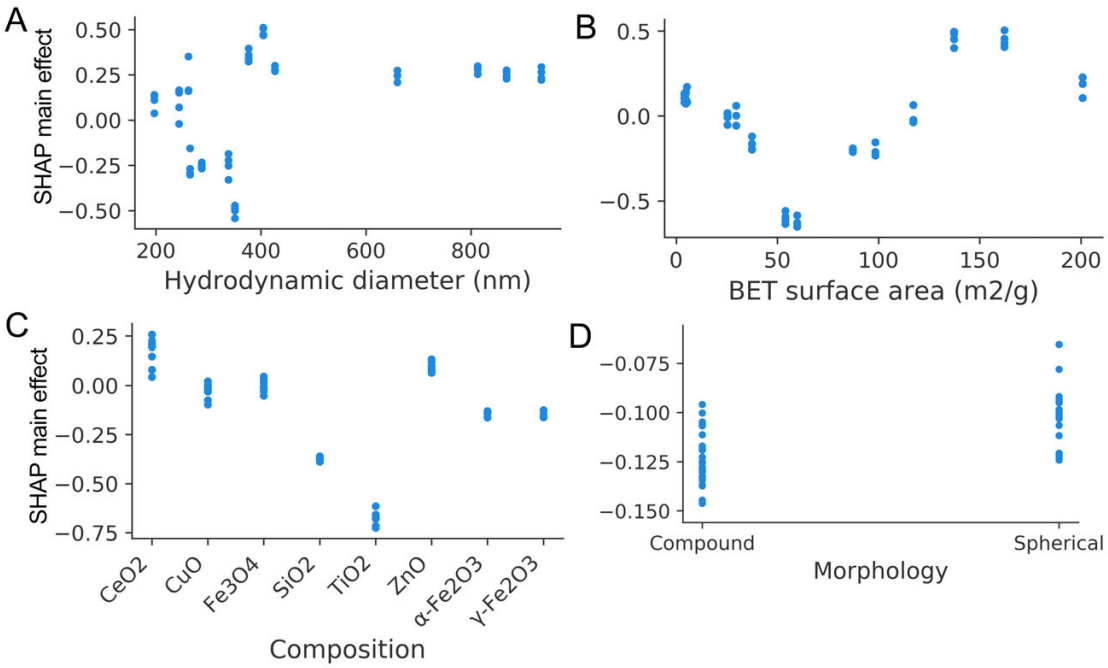

Fig. S17. SHAP main effects of hydrodynamic diameter (A), BET surface area (B), composition (C), and morphology (D).

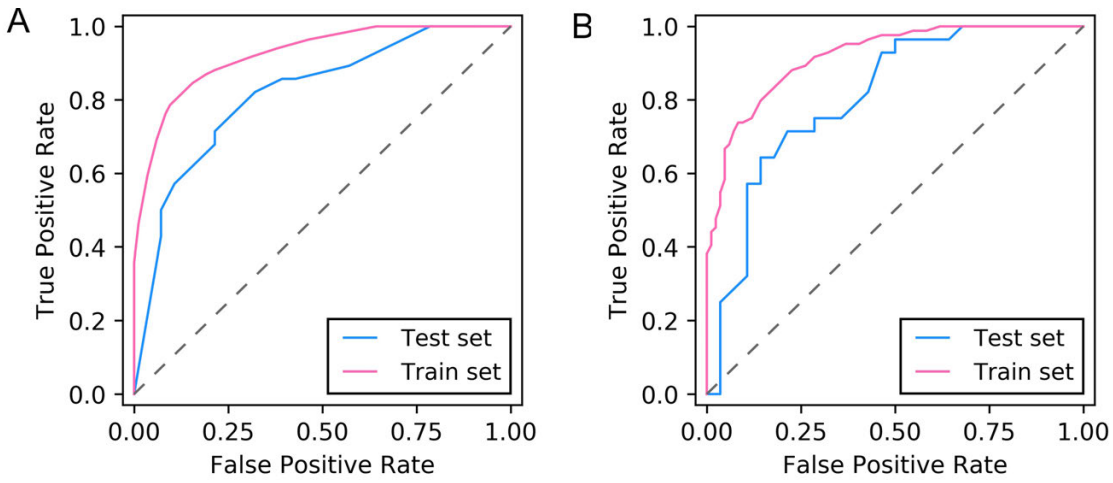

Fig. S18. ROC curve of the decision tree (A) and RuleFit (B) models trained on the fifth dataset split.

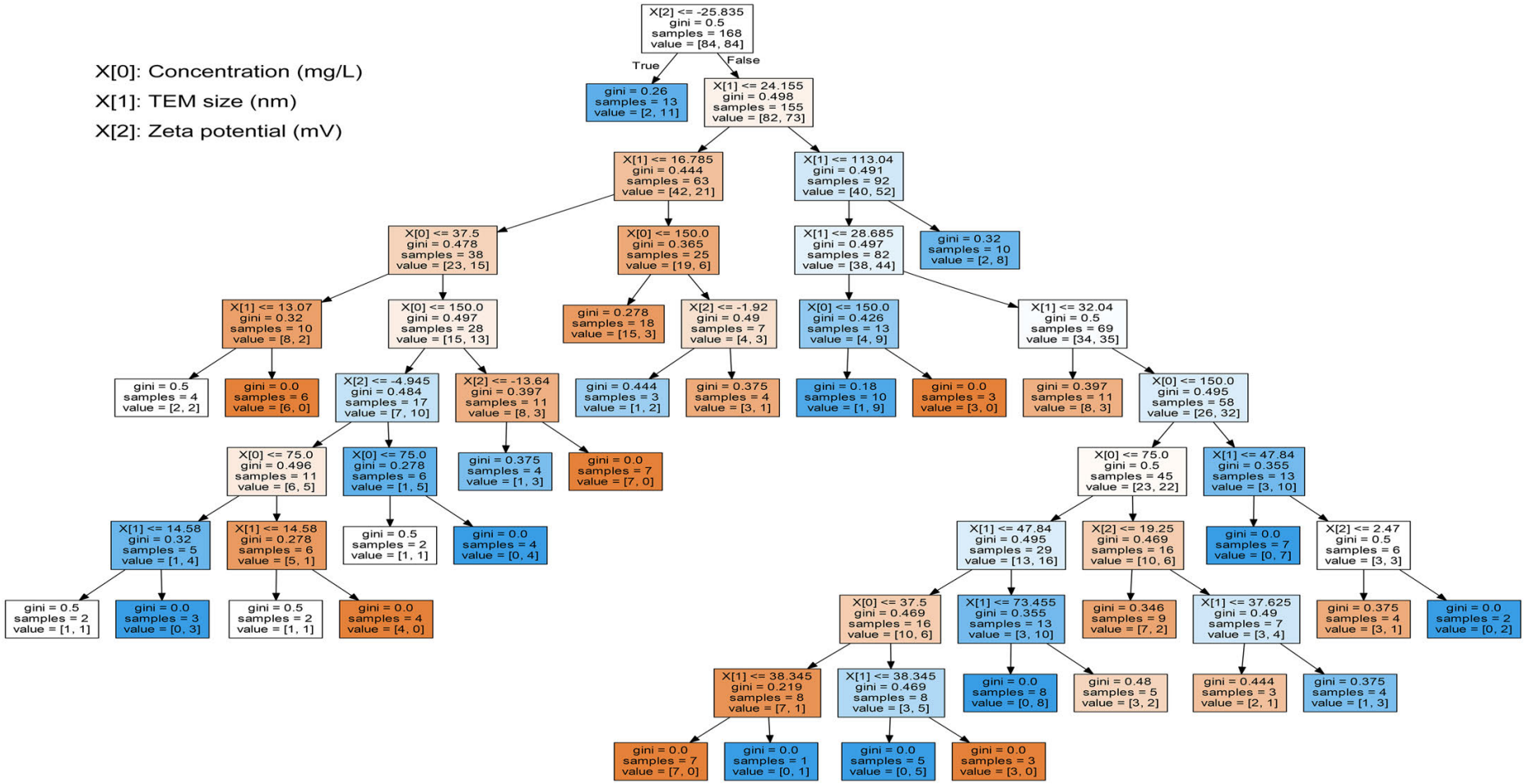

Fig. S19. The decision tree structure for root dry weight prediction based on three important features identified by post hoc interpretation of LightGBM models.

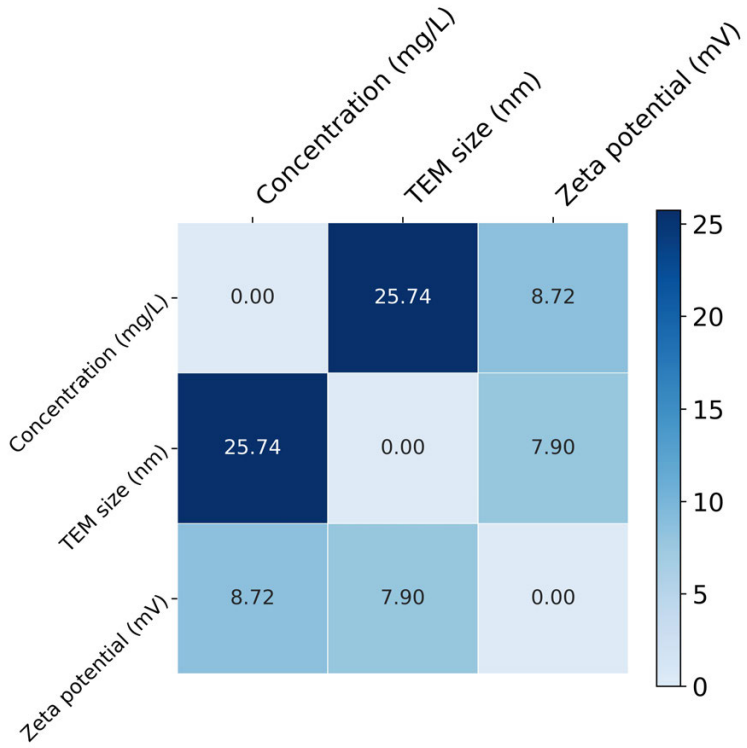

Fig. S20. SHAP interaction values of all features in the decision tree model.

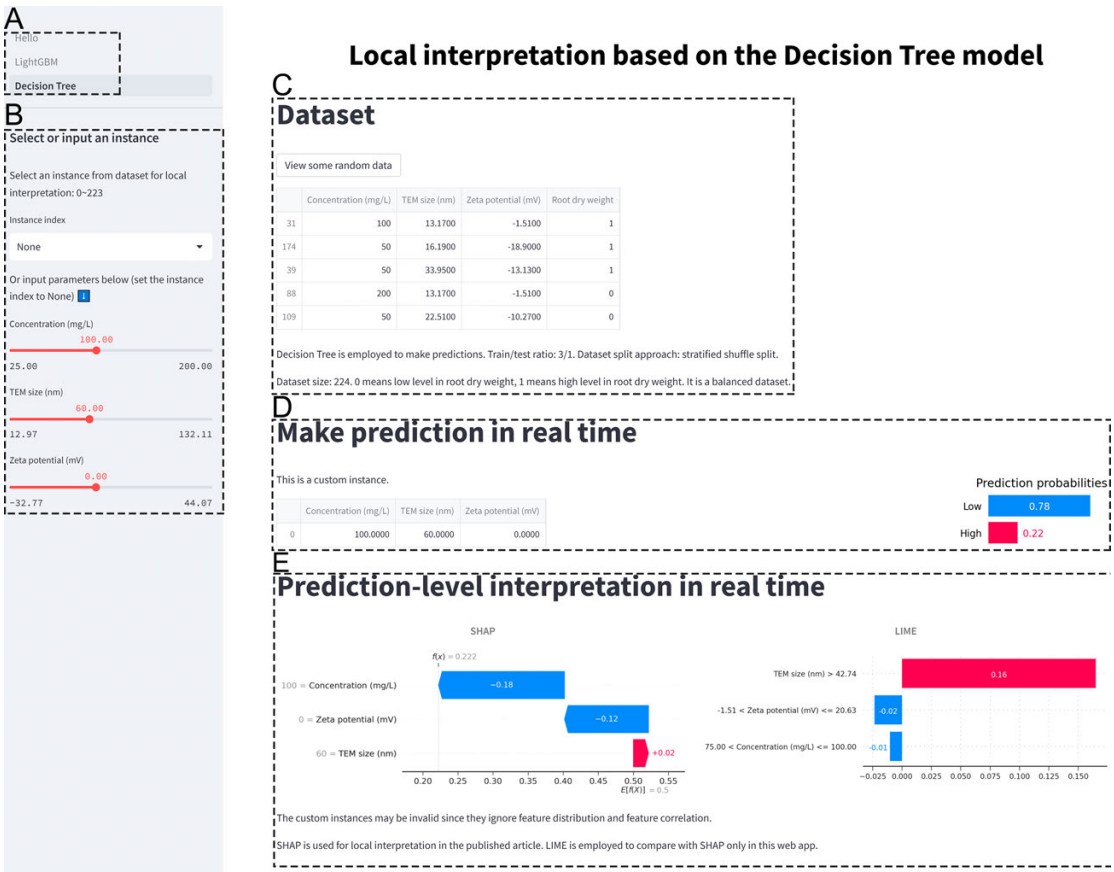

Fig. S21. The online interactive website for prediction-level interpretation (https://seed-nanopriming-isar.streamlit.app/). A, Navigation to different pages on this website. Hello: welcome page. LightGBM: local interpretation in the LightGBM model. Decision tree: local interpretation in the decision tree model. B, Select an instance from the used dataset or customize a sample. C, Dataset introduction and some random instances. D, Show the selected/custom instance and make a prediction. E, Prediction-level interpretation for this prediction.

Tables

Table S1.1. An overview of the used dataset and the prediction target (root dry weight).

Feature/target	Percentage/range
Composition	
SiO_2	14.29%
CuO	14.29%
CeO_2	14.29%
Fe_3O_4	14.29%
ZnO	14.29%
TiO_2	14.29%
α-Fe_2O_3	7.14%
γ-Fe_2O_3	7.14%
Morphology	
Compound	57.14%
Spherical	42.86%
TEM size	12.97~132.11 nm
TEM size SD	2.65~32.82 nm
Concentration	25, 50, 100, 200 mg/L
Hydrodynamic diameter	197~933.73 nm
PdI	0.17~0.81
Zeta potential	-32.77~44.07 mV
BET surface area	4.07~200.84 m2/g
Target: Root dry weight	1 for high level, 0 for low level

Table S1.2. Detailed analysis of the numerical features collected in this study.

	TEM size (nm)	TEM size SD (nm)	Concentration (mg/L)	Hydrodynamic diameter (nm)	PdI	Zeta potential (mV)	BET surface area (m2/g)
count	224	224	224	224	224	224	224
mean	39.19571	9.621429	93.75	459.0236	0.369286	3.439286	73.07857
std	32.9445	8.734165	67.17389	241.9576	0.215105	20.58079	60.58844
min	12.97	2.65	25	197	0.17	-32.77	4.07
25%	17.38	3.11	43.75	264.87	0.22	-12.93	25.31
50%	28.685	6.89	75	363.47	0.26	1.595	56.85
75%	42.74	10.98	125	660	0.54	20.63	117.09
max	132.11	32.82	200	933.73	0.81	44.07	200.84

Table S2. The version of the main software and packages used in this study.

Software/packages	Version
Python	3.10.8
scikit-learn	1.1.2
shap	0.39.0
PDPbox	0.2.1
imodels	1.2.5
lime	0.2.0.1
scipy	1.7.3
numpy	1.21.5
streamlit	1.13.0
R	4.2.2
agricolae	1.3-5
ropls	1.30.0

Table S3. The determined model hyperparameters of LightGBM models.

Random state	min_data_in _leaf	min_sum_hessian_ in_leaf	max_ bin	max_de pth	num_lea ves	learning_ rate
1	10	1	5	6	9	0.060
2	10	1	8	3	4	0.060
3	7	1	9	4	5	0.067
4	7	1	6	4	6	0.049
5	1	3	14	4	5	0.081
6	15	3	15	4	6	0.064
7	17	1	15	5	7	0.100
8	1	3	15	6	9	0.100
9	1	3	10	6	9	0.141
10	14	2	6	4	5	0.074

Table S4. Three group division of 56 nanopriming treatments based on the SRI under salinity stress.

Group	Treatments (Composition: Size(nm) : Concentration(mg/L))	Average SRI
High (7)	ZnO:30:200, CeO_2:<100:200, SiO_2:20:50, CeO_2:20-50:100, SiO_2:50:50, Fe_3O_4:50:100, Fe_3O_4:20:100	11.39
Middle (16)	CeO_2:<100:50, α-Fe_2O_3:30:100, CuO:40:25, SiO_2:20:200, ZnO:50:200, SiO_2:50:100, SiO_2:50:200, α-Fe_2O_3:30:25, CeO_2:20-50:25, α-Fe_2O_3:30:200, ZnO:30:50, ZnO:50:50, CuO:40:50, CeO_2:20-50:50, ZnO:50:25, SiO_2:50:25	10.74
Low (33)	SiO_2:20:25, Fe_3O_4:50:200, CeO_2:<100:100, SiO_2:20:100, ZnO:30:25, CeO_2:20-50:200, CuO:40:200, Fe_3O_4:20:50, CuO:50-100:25, α-Fe_2O_3:30:50, Fe_3O_4:50:25, TiO_2:40:25, TiO_2:20:25, γ-Fe_2O_3:<50:100, CeO_2:<100:25, TiO_2:20:50, TiO_2:40:100, TiO_2:40:200, Fe_3O_4:20:200, CuO:40:100, γ-Fe_2O_3:<50:50, γ-Fe_2O_3:<50:25, TiO_2:20:200, CuO:50-100:50, CuO:50-100:200, ZnO:30:100, CuO:50-100:100, ZnO:50:100, TiO_2:40:50, Fe_3O_4:20:25, Fe_3O_4:50:50, γ-Fe_2O_3:<50:200, TiO_2:20:100	9.99

Table S5. The 7-fold cross-validation results of established PLS-DA and OPLS-DA models.

Ion mode	Type	R^2X(cum)	R^2Y(cum)	Q^2(cum)	Treatments
positive	PLS-DA	0.679	0.994	0.912	SN_vs_SC
positive	OPLS-DA	0.679	0.994	0.725	SN_vs_SC
positive	PLS-DA	0.85	0.99	0.895	HdN_vs_HdC
positive	OPLS-DA	0.85	0.99	0.85	HdN_vs_HdC
negative	PLS-DA	0.508	0.954	0.744	SN_vs_SC
negative	OPLS-DA	0.683	0.994	0.87	SN_vs_SC
negative	PLS-DA	0.769	0.978	0.773	HdN_vs_HdC
negative	OPLS-DA	0.769	0.978	0.899	HdN_vs_HdC

Table S6. The rules determined by the RuleFit algorithm.

No.	Rule	Type	Coefficient	Support	Importance
39	Concentration (mg/L) <= 75.0 and Concentration (mg/L) > 37.5 and Zeta potential (mV) <= 5.565	rule	2.2548	0.1310	0.7607
3	Concentration (mg/L) <= 37.5 and TEM size (nm) <= 38.345 and TEM size (nm) > 28.685	rule	-3.0894	0.0536	0.6956
24	Concentration (mg/L) > 150.0 and TEM size (nm) <= 38.345 and Zeta potential (mV) <= -9.325	rule	2.6014	0.0714	0.6700
4	Concentration (mg/L) > 150.0 and TEM size (nm) <= 32.04 and Zeta potential (mV) > -9.325	rule	-2.2350	0.0952	0.6561
9	Concentration (mg/L) <= 75.0 and TEM size (nm) <= 24.155 and Zeta potential (mV) > -14.585	rule	-1.7272	0.1488	0.6147
21	Concentration (mg/L) <= 37.5 and Zeta potential (mV) > -13.03	rule	1.4855	0.1964	0.5902
34	Concentration (mg/L) > 150.0 and TEM size (nm) > 32.04	rule	1.8115	0.0893	0.5166
25	Concentration (mg/L) > 37.5 and Zeta potential (mV) > -9.325	rule	0.9328	0.4643	0.4652
27	Concentration (mg/L) > 75.0 and TEM size (nm) <= 47.84 and TEM size (nm) > 24.155 and Zeta potential (mV) <= 26.17 and Zeta potential (mV) > -25.835	rule	1.4086	0.1071	0.4357
19	Concentration (mg/L) > 37.5 and TEM size (nm) <= 47.84 and Zeta potential (mV) <= 26.17 and Zeta potential (mV) > -25.835	rule	-0.8501	0.4286	0.4207
32	Concentration (mg/L) <= 75.0 and TEM size (nm) > 24.155 and Zeta potential (mV) <= 11.285	rule	1.3193	0.1131	0.4178
8	Concentration (mg/L) <= 37.5 and TEM size (nm) > 28.685 and Zeta potential (mV) > -9.325	rule	-1.6738	0.0655	0.4140
14	Zeta potential (mV) <= -9.325 and Zeta potential (mV) > -25.835	rule	-0.7219	0.3036	0.3319
18	Concentration (mg/L) > 37.5 and TEM size (nm) <= 19.155 and TEM size (nm) > 13.07	rule	0.8465	0.1786	0.3242
7	Concentration (mg/L) <= 37.5 and TEM size (nm) <= 19.155 and TEM size (nm) > 13.07	rule	-1.3037	0.0595	0.3085
16	Concentration (mg/L) > 75.0 and TEM size (nm) <= 24.155 and Zeta potential (mV) > -14.585	rule	0.8025	0.1429	0.2808
17	Concentration (mg/L) > 75.0 and TEM size (nm) <= 113.04	rule	-0.5303	0.4702	0.2647
12	Concentration (mg/L) > 37.5 and TEM size (nm) <= 32.04 and TEM size (nm) > 16.785 and Zeta potential (mV) > -25.835	rule	-0.6072	0.2143	0.2491
11	Concentration (mg/L) > 75.0 and TEM size (nm) >	rule	-0.7658	0.1012	0.2310

	47.84 and Zeta potential (mV) > -25.835				
6	Concentration (mg/L) > 150.0 and TEM size (nm) <= 16.785 and Zeta potential (mV) > -13.64	rule	-1.1142	0.0417	0.2226
10	Concentration (mg/L) <= 75.0 and Concentration (mg/L) > 37.5 and TEM size (nm) <= 113.04 and TEM size (nm) > 38.345	rule	-0.7300	0.0536	0.1644
37	Concentration (mg/L) <= 150.0 and TEM size (nm) <= 28.685 and TEM size (nm) > 24.155 and Zeta potential (mV) > -25.835	rule	0.6808	0.0595	0.1611
29	TEM size (nm) <= 21.72 and TEM size (nm) > 19.155	rule	0.5940	0.0774	0.1587
28	Concentration (mg/L) <= 150.0 and TEM size (nm) <= 32.04 and TEM size (nm) > 16.785 and Zeta potential (mV) > 26.17	rule	0.6206	0.0595	0.1468
40	Concentration (mg/L) > 37.5 and TEM size (nm) > 113.04 and Zeta potential (mV) <= 26.17	rule	0.5028	0.0476	0.1071
5	Concentration (mg/L) <= 37.5 and Zeta potential (mV) <= -13.03 and Zeta potential (mV) > -25.835	rule	-0.5116	0.0417	0.1022
15	Concentration (mg/L) <= 37.5 and TEM size (nm) <= 113.04	rule	0.2161	0.2500	0.0936
26	TEM size (nm) > 19.155 and Zeta potential (mV) <= -11.6	rule	0.1976	0.2262	0.0827
20	TEM size (nm) > 21.72 and Zeta potential (mV) <= 26.17	rule	-0.1366	0.4762	0.0682
22	Zeta potential (mV) <= 26.17 and Zeta potential (mV) > -25.835	rule	-0.1407	0.7798	0.0583
23	TEM size (nm) <= 16.785 and Zeta potential (mV) <= 26.17 and Zeta potential (mV) > -25.835	rule	-0.1050	0.2262	0.0439
30	Concentration (mg/L) <= 75.0 and TEM size (nm) > 32.04	rule	0.0508	0.2024	0.0204
31	Concentration (mg/L) > 37.5 and TEM size (nm) <= 32.04 and TEM size (nm) > 16.785 and Zeta potential (mV) <= -21.52	rule	0.0828	0.0536	0.0186
13	TEM size (nm) <= 32.76 and Zeta potential (mV) <= 26.17 and Zeta potential (mV) > -25.835	rule	-0.0362	0.4405	0.0180
38	Zeta potential (mV) <= -25.835	rule	0.0327	0.0774	0.0087
35	TEM size (nm) <= 113.04 and Zeta potential (mV) <= -25.835	rule	0.0296	0.0774	0.0079
33	TEM size (nm) <= 38.345 and Zeta potential (mV) <= -25.835	rule	0.0158	0.0774	0.0042
36	TEM size (nm) > 19.155 and Zeta potential (mV) <= -22.95	rule	0.0119	0.0774	0.0032