Electronic Supplementary Materials

For https://doi.org/10.1631/jzus.A2200297

A novel stacking-based ensemble learning model for drilling efficiency prediction in earth-rock excavation

Fei LV, Jia YU✉, Jun ZHANG, Peng YU, Da-wei TONG, Bin-ping WU

State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300350, China

✉ Jia YU, yujia@tju.edu.cn

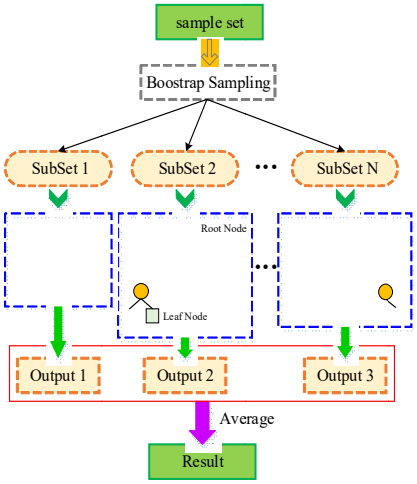

Fig. S1 The procedures of random forest (RF)

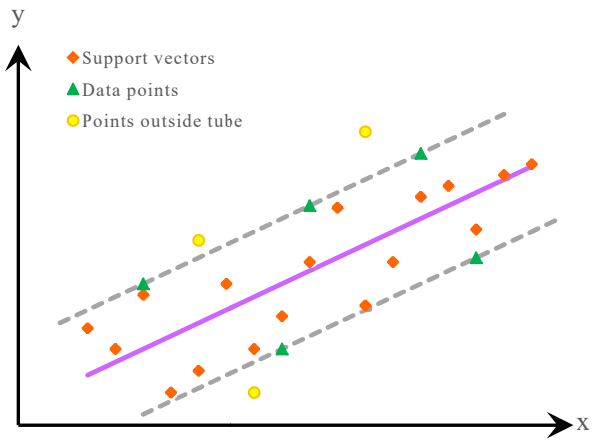

Fig. S2 Epsilon intensive band in support vector regression (SVR)

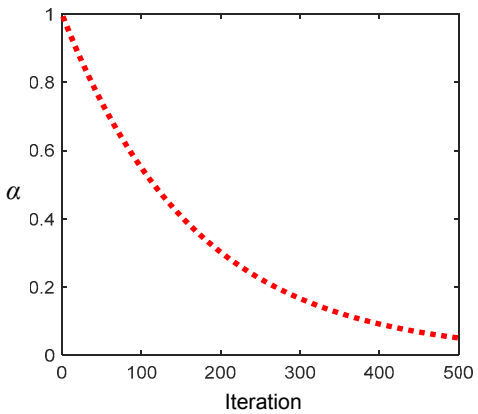

Fig. S3 Changes in α with growth in generations.

Table S1 Pseudocode of the improved cuckoo search optimization (ICSO) algorithm

Algorithm. The improved parameter optimization algorithm (CSO) for the prediction model

Initialize the parameters: the population size ps, G_{max}, α, α_{min}, α_{max}, probability pa, and the number of dimensions D;
Set the initial iteration number $t = 1$;
Randomly generate the initial populations of ps hosts' nests x_t^i (i =1, 2, …, ps);
The fitness function $f_t^i = f(X_t^i)$ (i =1, 2, …, ps, t =1, 2, …, G_{max});
while $t < G_{max}$ **do**
  Calculate the random step size α using equation (15) and (16);
  for $i \leftarrow 1$ to ps **do**
    Generate a new nest x_{t+1}^i making use of Levy flight in equation (10);
    Evaluate the fitness function of new solution $f_{t+1}^i = f(X_{t+1}^i)$;
    if $f_{t+1}^i < f_t^i$ **then**
      $X_{t+1}^i = X_t^i$;
      $f_{t+1}^i = f_t^i$;
    end
    else
      Preserve the solution of the t-th iteration;
    end
  end for
  Desert a fraction (pa) of worst solutions (nest);
  Generate new solutions (nest) utilising equation (14);
  Update and rank the solutions;
  Set $t = t + 1$;
end while
Output the optimal solution and corresponding fitness function value;

Table S2 Evaluation indicators of eight single ML models

Dataset	Model	MAPE	MAE	MSE	RMSE	R^2
Fold 1	BPNN	**0.140**	**0.249**	**0.186**	**0.432**	**0.778**
	XGBoost	**0.108**	**0.215**	**0.154**	**0.393**	**0.816**
	RF	**0.125**	**0.229**	**0.202**	**0.449**	**0.759**
	SVR	0.138	0.278	0.242	0.492	0.712
	LM	0.185	0.319	0.238	0.488	0.716
	GBDT	0.486	0.499	0.422	0.650	0.496
	ELM	0.842	1.293	2.721	1.650	-2.245
	DBN	0.853	0.713	0.840	0.916	-0.002
Fold 2	BPNN	0.145	**0.245**	**0.099**	**0.314**	**0.839**
	XGBoost	**0.107**	**0.188**	**0.079**	**0.281**	**0.872**
	RF	**0.138**	0.247	**0.091**	**0.302**	**0.852**
	SVR	**0.110**	**0.207**	0.108	0.329	0.824
	LM	0.166	0.308	0.188	0.434	0.693
	GBDT	0.377	0.401	0.241	0.491	0.607
	ELM	0.824	1.233	2.232	1.494	-2.631
	DBN	0.723	0.637	0.635	0.797	-0.032
Fold 3	BPNN	**0.190**	**0.216**	**0.121**	**0.348**	**0.830**
	XGBoost	**0.146**	**0.166**	**0.061**	**0.248**	**0.914**
	RF	**0.128**	**0.159**	**0.041**	**0.202**	**0.943**
	SVR	0.226	0.282	0.149	0.385	0.791
	LM	0.267	0.330	0.154	0.393	0.783
	GBDT	0.621	0.482	0.320	0.566	0.549
	ELM	0.682	0.873	1.298	1.139	-0.829
	DBN	1.141	0.799	0.922	0.960	-0.298
Fold 4	BPNN	**0.203**	**0.348**	0.358	0.598	0.666
	XGBoost	**0.157**	**0.288**	**0.297**	**0.545**	**0.723**
	RF	**0.183**	**0.327**	**0.316**	**0.562**	**0.705**
	SVR	0.204	0.355	0.375	0.612	0.650
	LM	0.262	0.380	**0.353**	**0.594**	**0.670**
	GBDT	0.601	0.591	0.603	0.776	0.437
	ELM	0.968	1.341	2.668	1.633	-1.494
	DBN	1.042	0.839	1.074	1.036	-0.004
Fold 5	BPNN	**0.212**	**0.220**	**0.087**	**0.296**	**0.898**
	XGBoost	**0.179**	**0.206**	**0.094**	**0.307**	**0.891**
	RF	**0.130**	**0.176**	**0.069**	**0.262**	**0.920**
	SVR	0.270	0.293	0.180	0.424	0.790
	LM	0.324	0.339	0.205	0.452	0.762
	GBDT	0.845	0.562	0.401	0.633	0.534
	ELM	1.115	1.061	1.473	1.214	-0.714
	DBN	1.550	0.903	1.108	1.053	-0.289

Note: the values in bold represent the top three best results of the evaluation indicators in each fold.

Table S3 Mean and Standard deviation of the evaluation indicators after 5-fold cross validation

	Model	MAPE	MAE	MSE	RMSE	R^2
Mean	BPNN	**0.178**	**0.256**	**0.170**	**0.398**	**0.802**
	XGBoost	**0.139**	**0.213**	**0.137**	**0.355**	**0.843**
	RF	**0.141**	**0.228**	**0.144**	**0.355**	**0.836**
	SVR	0.190	0.283	0.211	0.448	0.753
	LM	0.241	0.335	0.228	0.472	0.725
	GBDT	0.586	0.507	0.397	0.623	0.525
	ELM	0.886	1.160	2.078	1.426	-1.583
	DBN	1.062	0.778	0.916	0.952	-0.125
Standard deviation	BPNN	0.033	0.054	0.112	0.124	0.087
	XGBoost	0.031	0.046	0.096	0.119	0.076
	RF	0.024	0.066	0.114	0.147	0.102
	SVR	0.065	0.053	0.104	0.109	0.071
	LM	0.065	0.028	0.076	0.076	0.047
	GBDT	0.175	0.074	0.135	0.106	0.063
	ELM	0.163	0.192	0.663	0.237	0.847
	DBN	0.318	0.105	0.191	0.103	0.154
Coefficient of variation	BPNN	0.188	0.210	0.656	0.311	0.109
	XGBoost	0.226	0.217	0.701	0.336	0.091
	RF	0.171	0.292	0.793	0.414	0.122
	SVR	0.344	0.186	0.493	0.243	0.094
	LM	0.269	0.082	0.336	0.161	0.065
	GBDT	0.298	0.146	0.341	0.170	0.120
	ELM	0.184	0.166	0.319	0.166	-0.535
	DBN	0.299	0.135	0.209	0.108	-1.234

Table S4 Evaluation indicators of different single ML models as meta-learner

Dataset	Model	MAPE	MAE	MSE	RMSE	R^2
Fold 1	BPNN	0.125	0.151	0.041	0.202	0.935
	SVR	0.129	0.147	**0.039**	**0.198**	**0.938**
	XGBoost	0.138	0.195	0.068	0.261	0.892
	LM	**0.117**	**0.142**	0.039	0.198	0.938
	RF	0.153	0.189	0.065	0.256	0.896
	GBDT	0.471	0.349	0.202	0.450	0.677
Fold 2	BPNN	0.139	0.239	0.103	0.321	0.830
	SVR	**0.115**	**0.211**	**0.077**	**0.277**	**0.873**
	XGBoost	0.124	0.256	0.116	0.340	0.809
	LM	0.144	0.239	0.106	0.325	0.825
	RF	0.151	0.272	0.154	0.393	0.745
	GBDT	0.263	0.404	0.242	0.492	0.599
Fold 3	BPNN	0.149	0.257	0.187	0.433	0.825
	SVR	**0.135**	**0.236**	0.181	0.425	0.831
	XGBoost	0.197	0.310	0.248	0.498	0.769
	LM	0.149	0.251	**0.175**	**0.419**	**0.837**
	RF	0.159	0.251	0.223	0.473	0.792
	GBDT	0.816	0.558	0.475	0.689	0.557
Fold 4	BPNN	0.134	0.252	0.172	0.415	0.781
	SVR	**0.127**	**0.239**	**0.170**	**0.412**	**0.785**
	XGBoost	0.143	0.287	0.214	0.462	0.729
	LM	0.138	0.261	0.180	0.425	0.771
	RF	0.158	0.291	0.280	0.529	0.644
	GBDT	0.304	0.386	0.314	0.560	0.602
Fold 5	BPNN	0.148	0.203	0.192	0.438	0.809
	SVR	0.150	0.210	**0.183**	**0.428**	**0.818**
	XGBoost	0.220	0.289	0.248	0.498	0.754
	LM	**0.133**	**0.204**	0.194	0.441	0.807
	RF	0.236	0.288	0.222	0.471	0.780
	GBDT	0.851	0.527	0.461	0.679	0.542

Note: the values in bold represent the best results of the evaluation indicators in each fold.

Table S5 Mean and Standard deviation of the evaluation indicators after 5-fold cross validation

	Model	MAPE	MAE	MSE	RMSE	R^2
Mean	BPNN	0.139	0.22	0.139	0.362	0.836
	SVR	**0.131**	**0.209**	**0.13**	**0.348**	**0.849**
	XGBoost	0.164	0.267	0.179	0.412	0.79
	LM	0.136	0.219	0.139	0.361	0.835
	RF	0.171	0.258	0.189	0.424	0.771
	GBDT	0.541	0.445	0.339	0.574	0.595
Standard deviation	BPNN	0.010	0.044	0.065	0.101	0.059
	SVR	0.013	0.037	0.067	0.105	0.059
	XGBoost	0.042	0.045	0.082	0.106	0.064
	LM	0.012	0.048	0.065	0.102	0.062
	RF	0.036	0.042	0.082	0.106	0.091
	GBDT	0.278	0.092	0.125	0.108	0.053
Coefficient of variation	BPNN	0.072	0.201	0.470	0.279	0.070
	SVR	0.098	0.177	0.517	0.301	0.069
	XGBoost	0.254	0.168	0.459	0.258	0.081
	LM	0.091	0.221	0.470	0.283	0.075
	RF	0.212	0.162	0.436	0.250	0.118
	GBDT	0.515	0.207	0.368	0.188	0.088

S1 Benchmark functions to test the performance of ICSO

There are a few benchmark functions commonly used to test the performance and convergence speed of this optimization algorithm. In this study, these functions are given below.

Function 1: Bohachevsky function 1

The Bohachevsky functions all have the same similar bowl shape. The function formula is:

$$f(x) = x_1^2 + 2x_2^2 - 0.3\cos(3\pi x_1) - 0.4\cos(4\pi x_2) + 0.7 \tag{S1}$$

The functions are usually evaluated on the square $x_i \in$ [-100, 100], for all i = 1, 2. And its global minimum is

$$f(x^*) = 0, \text{ at } x^* = (0,0) \tag{S2}$$

Function 2: Bohachevsky function 2

The function formula is:

$$f(x) = x_1^2 + 2x_2^2 - 0.3\cos(3\pi x_1)\cos(4\pi x_2) + 0.3 \tag{S3}$$

The evaluation range of square xi and the global minimum is the same as Function 1.

Function 3: Bohachevsky function 3

The Bohachevsky functions all have the same similar bowl shape. The function formula is:

$$f(x) = x_1^2 + 2x_2^2 - 0.3\cos(3\pi x_1 + 4\pi x_2) + 0.3 \tag{S4}$$

The evaluation range of square xi and the global minimum is the same as Function 1.

Function 4: Beale function

The Beale function is multimodal, with sharp peaks at the corners of the input domain. The function formula is:

$$f(x) = (1.5 - x_1 + x_1x_2)^2 + (2.25 - x_1 + x_1x_2^2)^2 + (2.625 - x_1 + x_1x_2^3)^2 \tag{S5}$$

The function is usually evaluated on the square xi $\in$ [-4.5, 4.5], for all i = 1, 2. And its global minimum is

$$f(x^*) = 0, \text{ at } x^* = (3,0.5) \tag{S6}$$

The iterative curve of fitness optimization was shown in Fig. S1. It indicates that ICSO has faster convergence speed and better fitness search ability than CSO and PSO. And Table S1 shows that ICSO can find better fitness than CSO and PSO after 50 iterations (1.23 in Bohachevsky function 1, 1.04 in Bohachevsky function 2, 0.94 in Bohachevsky function 3, 0.40 in Beale function).

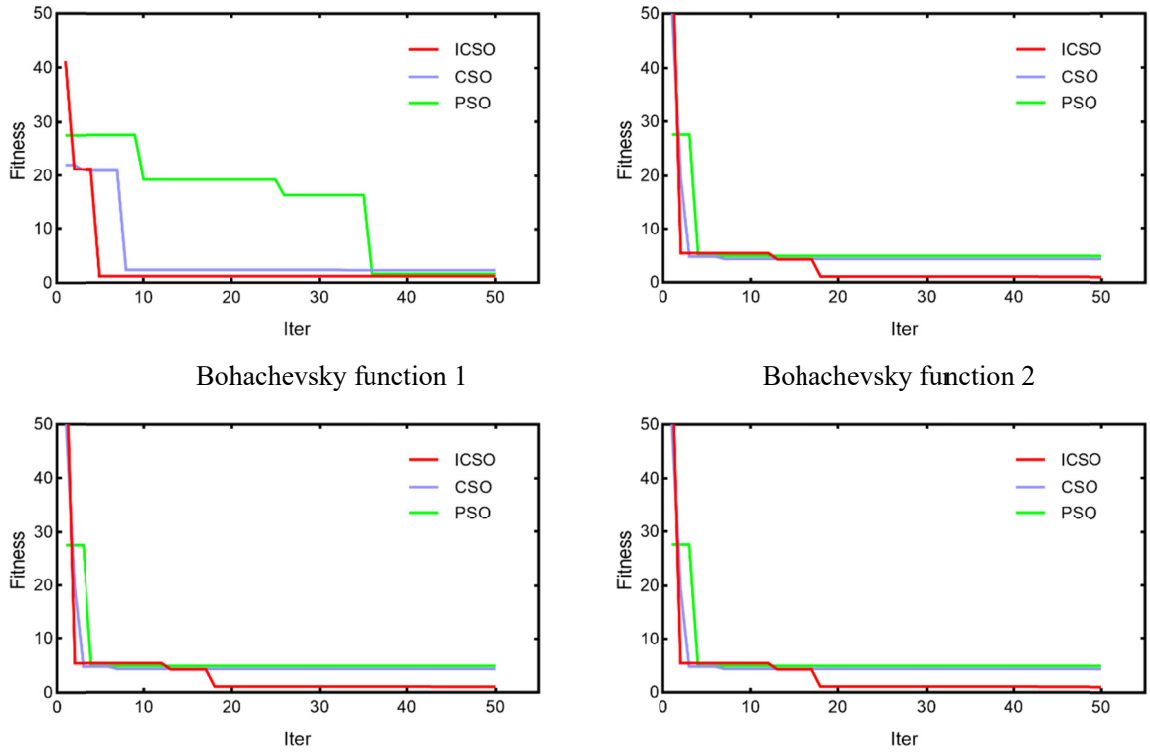

Fig. S4 The iterative curve of ICSO, CSO, and PSO using different benchmark functions

Table S6 The fitness value of different optimization algorithms after 50 iterations

Algorithm	Variables and fitness values	Bohachevsky function 1	Bohachevsky function 2	Bohachevsky function 3	Beale function
ICSO	x_1	0.794	-0.642	0.175	3.332
	x_2	-0.122	-0.204	-0.403	0.456
	Fitness	1.235	1.040	0.944	0.404
CSO	x_1	-0.578	0.714	-1.328	3.574
	x_2	-0.754	-1.293	-0.005	0.716
	Fitness	2.370	4.387	1.764	0.619
PSO	x_1	0.811	0.432	-0.722	2.372
	x_2	0.189	-1.466	-0.750	0.892
	Fitness	1.656	4.949	2.207	8.421