MULTIPLE PARAMETER OPTIMIZATION OF A LICORICE HARVESTER BASED ON ENSEMBLE MACHINE LEARNING AND IMPROVED GENETIC ALGORITHM

/
基于集成机器学习和改进遗传算法实现了甘草收获机的多参数优化

Jinyu SONG, Yonglei LI*), Xiaopei ZHENG, Lipengcheng WAN, Zongtian LIU
China Agricultural University, College of Engineering, Beijing / China
Tel: +86-13811692767; E-mail: liyl0393@163.com;
Corresponding author: Yonglei Li

ABSTRACT

Optimizing parameters is a crucial step in designing mechanical structures and a primary means of raising equipment efficiency. This paper proposes a multi-parameter optimization technique that combines an improved genetic algorithm (IGA) and ensemble machine learning (EML) to optimize a licorice harvester's work and structure parameters. The EML model is trained using a small sample dataset built on the coupled DEM-MBD (Multi-body Dynamics Coupled Discrete Element Method) simulation model. The impact of base learner diversity and quantity on the model's prediction accuracy is investigated. Using EML and IGA, the parameters of a licorice harvester are optimized. It is also contrasted with conventional response surface model (RSM) parameter optimization techniques. The study results show that the EML with KNN +lightGBM + catBoost as the base learner and linear as the meta-learner has an R^2 of 0.959, MAE of 0.048, and RMSE of 0.06. In comparison to the RSM, EML-IGA reduces resistance by 18.16% and specific power consumption by 21.33%; in comparison to the EML and Pre-improvement genetic algorithm(PIGA), it reduces resistance by 11.36% and specific power consumption by 11.19%. It provides a reference for intelligent parameter optimization methods.

摘要

参数优化是机械结构设计过程中必不可少的环节，也是提高机械工作效率的主要途径之一。本研究通过集成学习与改进遗传算法结合提出一种多参数优化方法对甘草收获机的结构和工作参数进行优化。基于 DEM-MBD 耦合仿真模型构建小样本数据集对集成学习模型进行训练，并探究基学习器的数量与多样性对集成学习模型预测精度的影响。利用集成学习结合改进遗传算法对甘草收获机的多个参数进行优化。并与传统的响应面参数优化方法进行对比。研究结果表明，以 KNN+lightGBM+catBoost 为基学习器，线性拟合为元学习器的集成学习模型，其 R^2 为 0.959，MAE 为 0.048，RMSE 为 0.06。其相较于改进前的遗传算法的优化结果，阻力降低 11.36%，比功耗降低 11.19%，相较于传统的响应面分析法，阻力降低 18.15%，比功耗降低 21.33%，为智能化参数优化方法提供参考。

INTRODUCTION

The main component of the licorice harvester is the digger device, whose work and structure parameters directly impact the machine's resistance and power consumption. Both domestically and internationally, researchers have conducted a great deal of study on the optimization of the parameters of the harvesting device for deep root crops to address the issues of high digging resistance and high power consumption in deep root crops. *Zhang et al., (2024)*, optimized three working parameters of a residual film recycler using RSM to improve its pick-up rate. *Awuah et al., (2022)*, optimized the parameters of the vibratory digging shovel based on DEM and RSM, significantly reducing the working resistance. The RSM utilized in the previously mentioned parameter optimization techniques is primarily useful for optimization variables fewer than or equal to 4. An excessive number of optimization variables might result in issues such as local optimal solutions and inaccurate fitting. As machine learning advances, more academics are using it to predict regression using multivariate inputs. *Huang Lvwen et al, (2023),* utilized 40 feature values as inputs into the LSTMED-MLP model to forecast the soluble solids content of apples. *Ge et al., (2023)*, trained eight machine learning algorithms by multiple input features with different importance. It can be seen that machine learning is suitable for the nonlinear fitting of multivariate inputs for prediction.

However, most of the research has generally focused on the predictive performance of individual machine-learning models. Single models' limited adaptability makes them frequently unable to handle complex problems completely *(Cao et al., 2024)*.

On the other hand, by combining the benefits of several models, EML can successfully enhance the generalization performance of prediction models and lower prediction errors. *Liu Tan et al, (2024),* used EML to estimate the photosynthetic rate of greenhouse tomatoes, which increased the prediction's accuracy and stability. *Zongquan, (1997),* researched the predictive accuracy of Stacking's EML for cracking in reinforced concrete against that of a single machine learning model; it was discovered that the model performed better. Consequently, it is now worthwhile to research how to fully utilize the benefits of a single learner in order to enhance the model's generalization performance and prediction accuracy. Researchers have started combining machine learning and simulation modeling techniques to solve various real-world problems since numerical simulation techniques have become more popular. *Yu et al.,(2023),* trained BP neural networks using datasets obtained from numerical simulations. *Liao et al., (2021),* combined DEM and deep learning methods to predict particle flow behavior in a wedge-shaped hopper.

An intelligent optimization technique, genetic algorithm (GA), is utilized to find the objective function's optimal solution *(Aote et al., 2023)*. Although it has a straightforward structure, it has limitations on local search capability and population variety. A few academics started enhancing GA. *NING Fanghua et al, (2024),* combined the NEH and random generation to produce high-quality first populations as an alternative to the conventional technique of randomly generating starting populations. *Wei et al, (2024),* introduced simulated annealing algorithm into the genetic algorithm, which improved the local search ability of the traditional genetic algorithm. The crossover and mutation probabilities of GA should be adjusted adaptively; however, this is rarely done by academics. Doing so would increase the genetic algorithm's capacity for global search and speed of convergence.

The EML model is trained using a small dataset built on the coupled DEM-MBD simulation model. The impact of base learner diversity and quantity on the model's prediction accuracy is investigated. EML-IGA was possible to predict licorice harvester performance indicators quickly and accurately. A uniform distribution of the initial population and adaptive adjustment of crossover and mutation probabilities are achieved by improving upon some of the shortcomings of conventional genetic algorithms. Subsequently, the optimization results of the response surface analysis method are compared with the EML - IGA to obtain a method suitable for complex multi-parameter optimization problems. The combination of structural and working parameters was also optimized to achieve the licorice harvester's minimum resistance and power consumption.

MATERIALS AND METHODS

Small sample dataset construction

- **Physical structure of the licorice harvester**

The structure of the licorice harvester is shown in Fig.1, which is mainly composed of digging device, suspension system, vibration system, excitation device and frame. Fig.2 shows the main structural parameters, digging inclination α, dispersal inclination β, dispersal length L, dispersal spacing D, and working width d. Working speed V, harvesting depth H, vibration frequency f, and crank amplitude A are the primary operating parameters.
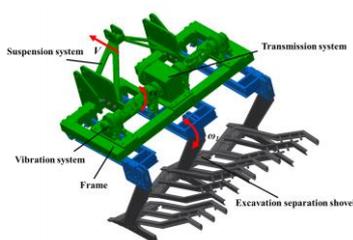
| Fig. 1 - Physical structure of the licorice harvester | Fig. 2 - Main structural parameters of excavating shovel | Fig. 3 - Multi-body dynamics simulation model of the licorice harvester |
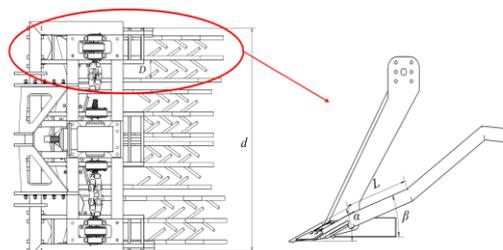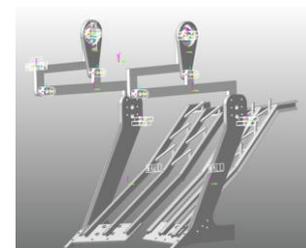
● **MBD Simulation Model Building**

The licorice harvester is first structurally simplified. Parts of the frame, gearbox, vibration system, etc. that do not affect the simulated motion setup are removed, and only half of the digging shovel structure of the original unit is retained. It accomplishes the objective of reducing the simulation time for model coupling and ensuring that the model moves by the working principle. The working width of the simplified unit is 840 mm. Fig. 3 shows the multi-body dynamics simulation model of the licorice harvester. Import the simplified 3D model into RecurDyn in .sat format and add motion and drives between parts. Two drives in total are added to the model. The first adds a forward drive with the drive function *V*time* to the frame for linear motion. The other adds a vibration drive to the eccentric block for rotational motion with a drive function of $\frac{f*360}{57.3°} * time$ (*time* is the total simulation time).

● **DEM simulation model construction**

The soil's calibration and the soil particle interaction model selection are crucial to producing a DEM model. The soil samples calibrated were brown loam soils in Beijing. The density of the soil was 2130 kg/m^3, shear modulus 0.96 MPa, Poisson's ratio 0.36 *(Song Jiannong et al, 2021)*. The soil has a sticky texture and a high degree of bonding between soil particles. The Hertz-Mindlin with JKR Cohesion contact model was selected for parameter calibration because it takes the influence of bonding force into account and works well with cohesive soils with high adhesion forces *(Junwei et al., 2019)*. The static repose angle of the soil served as an indicator for calibrating the target soil. The actual soil repose angle test was repeated five times and then averaged to obtain the actual soil repose angle θ = 31.42°. The natural soil static repose angle values were obtained by processing the raw photos of the static repose angle test using PyCharm software, which allowed for determining the soil repose angle. The processing is shown in Fig. 4.

Using soil-soil collision recovery coefficient *X1*, soil-soil static friction factor *X2*, soil-soil rolling friction factor *X3*, and JKR surface energy *X4* as the test factors and soil repose angle as the test index, Box-Behnken simulation test was conducted to determine soil-soil contact parameters. The DEM model for the calibration test is shown in Fig. 5. The soil particles were defined as spherical, measuring 6 mm in size, and their distribution followed a conventional normal distribution with a 0.05 variance. Table 1 displays the range of JKR surface energies as well as the range of soil-soil contact coefficients needed for calibration, together with the organic glass intrinsic parameters needed for the tests and the soil-organic glass contact coefficients that have been observed in the literature *(Fangping et al., 2020)*. The software Design-Expert 13 carried out the Box-Behnken simulation test program and optimized the results. Optimal parameter combinations of contact parameters are obtained: *X1=0.39, X2=0.4, X3=0.05, X4=7.22*. The combination of contact parameters is effective since the simulated value of the repose angle is 30.64°, and the relative error with the observed value is 2.48%, indicating the relative error is less than 5%. The digging shovel's material is 65Mn, its density is 7865 kg/m^3, its Poisson's ratio is 0.3, and its shear modulus is 7.9x107 MPa *(Song Jiannong et al, 2021)*. The contact parameters of soil and digging shovel are as follows: the collision recovery coefficient *e* is 0.5, the static friction factor μ_s is 0.3, and the dynamic friction coefficient μ_r is 0.1 *(Zhang et al., 2017)*.

For the excavation shovel to operate normally, the DEM simulation model needs to be configured as an earth slot since the shovel's real operating environment is a complete field. The length of the soil box was set at 3500 mm, the width at 950 mm, and the height at 800 mm, based on the excavation shovel's working width and speed.

(a) (b) (c) (d)
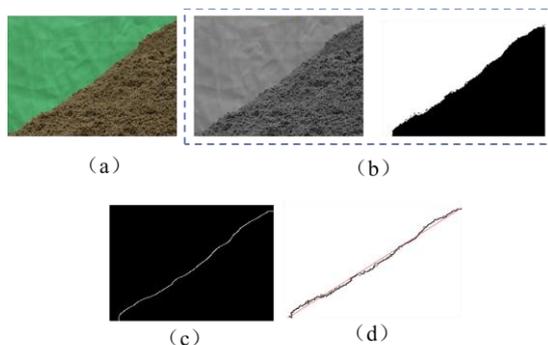
Fig. 4 - The process of soil repose angle processed by Pycharm
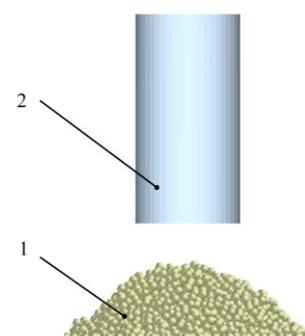
Fig. 5 - DEM model of calibration test

Table 1

Calibration model simulation parameters

Parameter	Value
Poisson's ratio of soil particles	0.36
Density of soil particles /(kg/m3)	2130
Shear modulus of soil particles /MPa	0.96
Poisson's ratio of organic glass	0.37
Density of organic glass/(kg/m3)	2.5×10^3
Shear modulus of organic glass /MPa	100
Soil-organic glass Recovery coefficient ball and steel	0.3
Soil-organic glass static friction coefficient<	0.41
Soil-organic glass rolling friction coefficient<	0.01
soil-soil collision recovery coefficient	0.35 ~ 0.7
Soil-soil static friction	0.4 ~ 1.1
Soil-soil rolling friction	0.05 ~ 0.25

- **Coupled DEM-MBD simulation model of licorice harvester**

The two digging shovels from the MBD simulation model are exported in WALL format and then loaded into the EDEM2020 software to establish the model. The time step of the MBD simulation was set to 15 samples per vibration cycle. In summary, the STEP design equation for the MBD model is *f*15*time*, where time is the overall simulation duration. The DEM's time step is 20%, and the save interval is 0.005 s.

- **Design of dataset construction methods**

This study proposes using a small sample-based simulation dataset for machine learning training to increase design efficiency. The resistance F_q and the power consumption P_{wk} (power consumption per unit volume of soil handled by the licorice harvester) that the harvester experiences while in operation were utilized as optimization indications. F_q is extracted in the DEM model simulation results.

The driving torque T_q is extracted directly by the result function in RecurDyn. P_{wk} was calculated using Equation (1).

$$P_{Wk} = \frac{\overline{F_q}}{1000DBt} + \frac{\overline{T_q}R}{9550VDBt} \tag{1}$$

where:

$\bar{F}_q$ is the average of resistance in the effective range, N;
$\bar{T}_q$ is the average of driving torque with the value greater than 0, N.m;
R is the crank speed, r/min;
t is the working time, s;
V is the forward speed, m/s;
B is the working width, mm;
D is the digging depth, mm.

However, F_q and P_{wk} are calculated as a combined performance index Z utilizing linear weighting in accordance with the Entropy technique in order to simplify the genetic algorithm optimization. The entropy method is more objective, assigning weights based on sample data sets *(de Blas et al., 2021)*. According to the calculations, the weight of P_{wk} is 0.757, and the weight of F_q is 0.243. The licorice harvester's primary structure and operating parameters have been described above. In this study, the parameters significantly affecting the composite indicator Z were screened by the Plackett-Burman test. Six design factors significantly impacted the composite index Z: vibration frequency f, crank amplitude A, harvesting depth H, operating speed V, digging inclination angle α, and dispersal inclination angle β. To summarize, this study generates a small sample dataset by utilizing the composite index Z as the optimization index and the licorice harvester's six structural and operating characteristics as the optimization variables. The dataset was produced by a 6-factor Box-Behnken test with 54 groups, created with the software Design-Expert 13.

Ensemble Machine Learning Model Building
- **Building Methods for Ensemble Machine Learning**

Ensemble Machine Learning (EML) is one of the more advanced machine learning paradigms *(Ribeiro et al., 2022)*. The main principle is to use one of the three strategies—Boosting, Bagging, or Stacking—to aggregate the prediction result of several learners (base learners) to achieve more excellent prediction performance than a single learner. Stacking has the flexibility to take full advantage of different base learners, as well as the ability to select meta-learners based on specific problems and samples. Therefore, a stacking integration strategy was selected for this study. There are two layers in the stacking integration model: Level 1, which comprises several base learners, each of which produces individual predictions. The meta-estimator, present in Level 2, used inputs of predictions from several base learners in Level 1 to learn more and generate the final integrated predictions. The base learner selects the eight machine learning models—Random Forest (RF), Decision Tree (DT), K Nearest Neighbors (KNN), Ridge Regression (RR), AdaBoost, lightGBM, catBoost, and XGBboost —that are commonly employed for regression prediction. The meta-learner selects a more straightforward linear regression to avoid overfitting and improve the model's generalization performance.

Evaluation index of Ensemble Machine Learning

The accuracy of the EML model is assessed in this work using three indices: the root mean square error (RMSE), the mean absolute error (MAE), and the coefficient of determination (R^2) *(Wu et al., 2022)*. The model's quality of fit is shown by the R^2, which has a value range of 0 to 1. The closer the value is to 1, the better the model. The MAE and RMSE values represent the model's prediction error; the lower the number, the smaller the prediction error and the higher the model's fitting accuracy.

Improvement methods for genetic algorithms
- **Improvements in initialization of populations**

The population of the traditional genetic algorithm is not uniformly distributed and is initialized with a random distribution within the optimization range of the independent variables. This could lead to the algorithm finding a local optimum solution *(Qiao et al., 2022)*. Therefore, rather than using random distribution, Circle Chaotic Mapping is used in this work. Circle mapping can improve the algorithm's capacity for global search and offer a proper exploratory mechanism. The mapping form is indicated in Equation (2).

$$D_{i+1} = mod(D_i + a - \frac{b}{2\pi}\sin(2\pi D_i), 1) \tag{2}$$

where: D_i denotes the current mapping state value, located in the interval (0,1); D_{i+1} is the next state value. a and b are the control parameters, which usually take the values of $a=0.5$ and $b=0.2$. The mod function guarantees that the mapping's outcome stays inside (0, 1).

- **Adaptive improvement for crossover and mutation probability**

The Gompertz function is a mathematical model that Gompertz initially proposed. It is typically used to represent a system or process that increases in speed at the rate of the first fast change rule followed by a slow one *(Yin et al., 2021)*. The crossover, mutation probability, and function curve change rules are similar. Consequently, this study refines the Gompertz function to create an adaptive adjustment formula for the crossover and mutation probabilities. The general form of the Gompertz function is shown in Equation (3). An evolutionary coefficient R must be proposed to characterize the population's degree of evolution to satisfy the requirement that the probability of crossover mutation is by the population's degree of evolution. Individual fitness values are small and discrete during the early stages of genetic algorithm population evolution. However, as the population ages and approaches the ideal solution, its fitness values become more concentrated. The notions of expectation and variance are introduced to represent the change in fitness values. The population fitness value increases throughout the evolutionary process while the variance decreases. Therefore, the evolutionary coefficient R expression can be designed as Equation (4). To satisfy the adaptive features, the general form of the Gompertz function was combined with the evolutionary coefficients R to form the crossover and mutation probability adaptive adjustment formulas in Equation (5) and (6), where the *a, b*, and *c* coefficients are used to set the range of crossover mutation probability values and their trends.

$$f(x) = ae^{-be^{-cx}} \tag{3}$$

$$R = \frac{EX+1}{\sqrt{DX}} \tag{4}$$

$$f(x) = 0.9e^{-0.05e^{0.4R}} \tag{5}$$

$$f(x) = 0.1e^{-0.05e^{0.8R}} \tag{6}$$

where: x is the independent variable; a, b, c are positive real parameters. a denotes the maximum value; b controls the initial growth rate; and c controls the steepness of the growth curve. P_c is the crossover probability, ranging from $(0,0.9)$; P_m is the variation probability, ranging from $(0,0.1)$; and R is the evolutionary coefficient, ranging from $(0, +\infty)$.

Design of multi-parameter optimization methods

Three optimizing strategies will be developed to optimize the licorice harvester's six structures and operating parameters. The first is based on traditional response surface analysis; the second is EML combined with a genetic algorithm for optimization; and the third is EML combined with the improved genetic algorithm for multi-parameter optimization. All three methods are based on the same dataset.

RESULTS
Exploration of the influence of structural working parameters on Z

To guide the actual design and optimization, the impact of operational and structure parameters on the performance indices was examined independently. Fig 6 shows the pattern of influence of single factors on Z. The remaining factors are at intermediate levels when examining the pattern of influence of individual components.
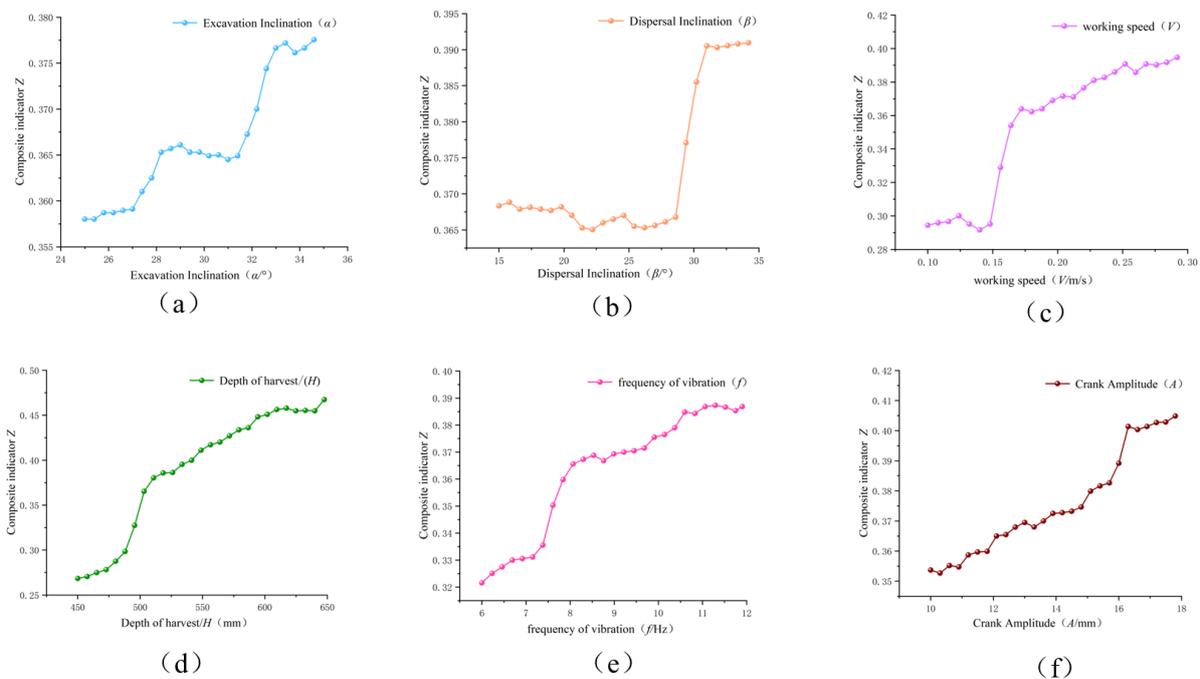
Fig. 6 - Single-factor influence pattern diagram
(a) Laws of influence of α on Z; (b) Laws of influence of β on Z; (c) Laws of influence of V on Z;
(d) Laws of influence of H on Z; (e)Laws of influence of f on Z; (f) Laws of influence of A on Z

Fig. 6(a) shows the pattern of excavation inclination angle on *Z*. As α increases, *Z* is gradually increasing. However, when *α* is at 28-31°, *Z* is unchanged and tends to decrease slowly. Beyond 32°, *Z* increases rapidly before stabilizing. Thus, resistance and power consumption might quickly surge when the digging inclination is too high. Fig. 6(b) shows the law of β on *Z*. *Z* is nearly constant between 15 and 29° of *β*; however, above 29°, *Z* climbs quickly to a higher level and then stays nearly constant. Consequently, the design of β should be smaller than 29°. Fig. 6(c) is the law of the effect of *V* on *Z*. When *V* exceeds 0.15 m/s, *Z* climbs quickly before rising slowly. In general, resistance and power consumption rise with *V*. Consequently, the deep-rooted crop harvester's working speed should be limited to roughly 0.15 m/s. The impact of *H* on Z is shown in Fig. 6(d). Resistance and power consumption rise with increasing *H*. 500 mm of *H* causes a spike in resistance and power consumption.

Therefore, not only will resistance and power consumption increase with increasing H (more than 500 mm), but the operation will also become increasingly challenging. (e) is the law of the effect of f on Z. There is an increase in power consumption with increasing f. Power consumption rises in proportion to the f. There is a spike in power consumption at 7.5 Hz. However, the power consumption is nearly constant and somewhat steady when the frequency is at 8-10 Hz. (f) is the law of A on Z. As the A increases slowly, the power consumption increases slowly.

In conclusion, there is a positive correlation between each factor's laws of impact over resistance and power consumption. It is important to keep each aspect within a narrow range when designing.

Comparative evaluation of different combinatorial basis learners' prediction effects for ensemble machine learning

The choice of base learners is the most crucial step in training ensemble machine learning models. This section investigates how the quantity and kind of base learners affect the model's prediction accuracy. Alternatives for the base learner, RF, DT, KNN, RR, AdaBoost, lightGBM, catBoost, and XGBboost, have been selected. Fig X displays the prediction results received after the model has been trained. As the Fig. 7 illustrates, all five of the prediction models—AdaBoost, lightGBM, catBoost, XGBboost, and RF— have good robustness, with R^2 values over 0.9 and RMSE and MAE errors hovering around 0.08, indicating excellent prediction accuracy. By comparison, R^2 was less than 0.8, and MAE and RMSE errors were more significant than 0.12, indicating lower prediction accuracy for KNN and ridge regression compared to the other models. The more accurate base learners—AdaBoost, lightGBM, catBoost, XGBboost, and RF—are chosen in order to guarantee the predictive capacity of ensemble machine learning. Because unlike the five model-building procedures mentioned above, KNN was also used as the primary learner to ensure the diversity of the underlying learners.
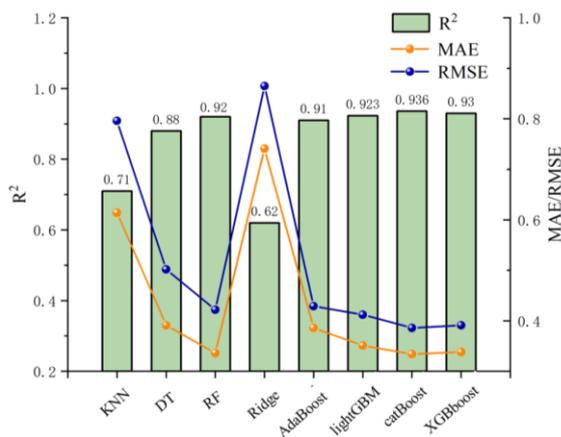
Fig. 7 - Base Learner Prediction Accuracy Graph   **Fig. 8 - Robust graph of prediction accuracy**
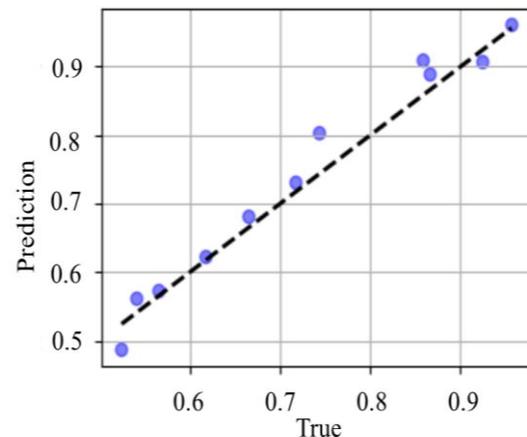
This study will use two-by-two, three-by-three, four-by-four, five-by-five, and six-by-six combinations to systematically analyze the impact of the set of base learners in different configurations on the model performance in order to investigate the impact of the number and diversity of base learners on the prediction accuracy of the integrated learning model. A linear model is used to fit the meta-learner in order to prevent the ensemble machine learning model from overfitting. Fig.9 displays the prediction accuracy of the ensemble machine learning using various base learner combinations. (a) displays the primary learner's two-by-two combination model's prediction performance. The model with the highest prediction accuracy is model 8-KNN+catBoost. The prediction performance of the three combined basic learner models is displayed in (b). Model 6-KNN+lightGBM+catBoost has the best robustness and lowest error. 0.959 is the R value, 0.048 is the MAE, and 0.06 is the RMSE. The prediction performance of the four combined basic learner models is displayed in (c). Models 4 and 5 both have an R^2 of 0.949, an MAE of 0.05, and an RMSE of 0.066, indicating higher accuracy and fewer prediction mistakes. The prediction performance of the five combined base learner models is displayed in (d). The three models have similar and strong prediction abilities. The error is approximately 0.066, and the R^2 approach is 0.95. The final model is a combination of six, and it has an RMSE of 0.068, an MAE of 0.052, and an R^2 of 0.939.

In conclusion, it was discovered that when the essential learners are merged two by two, the KNN+catBoost model performs best across all combinations. The models with three and four combinations have comparable predictive power, whereas those with six have less predictive power. Also, it was discovered that KNN models were included in the two-by-two model 8-KNN+catBoost, three-by-three model 6-KNN+lightGBM+catBoost, and four-by-four model 4-KNN+AdaBoost+lightGBM+catBoost combinations. Model KNN has the lowest predictive power R^2, at just 0.71, as the preceding section has shown. This means that in addition to models with high individual predictive ability, models of various types are also needed to select base learners. Ensemble machine learning's prediction accuracy rarely increases appreciably once the base learner count reaches a certain point. Thus, the option of base learners is 2 or 3, given the combination of the arithmetic cost and the need for model prediction accuracy.

Therefore, in this study, ensemble machine learning with KNN+lightGBM+catBoost, which has the best prediction accuracy, as the base learner and linear fitting as the meta-learner was chosen as the prediction model for the performance index of the licorice harvester. As shown in Fig. 8, the prediction robustness plot of this model indicates its strong prediction ability.

Comparative analysis of optimization capabilities of different methods

From the above, the ensemble machine learning model identified in this study has KNN+lightGBM+catBoost as the base learner in the first layer and linear regression as the meta-learner. Pre- and post-improved genetic algorithms will be coupled with the ensemble machine-learning model to optimize the licorice harvester's operating and structure parameters. *H* was fixed at 600 mm during the optimization process, and the other parameters were optimized. The pre-improved optimization algorithm optimized the parameter combinations as *α=26°, β=27°, V =0.1 m/s, f =6 Hz, A =11 mm*. the improved optimization algorithm optimized the parameter combinations as *α=25°, β=25°, V=0.1m/s, f=6Hz, A=11mm*. The findings of an ANOVA based on response surface analysis for Z optimization are displayed in Table 2. The response surface model's R^2 of 0.86 is less than the ensemble machine learning model's R^2 of 0.959, even though the response surface model is highly significant ($p < 0.01$) and the misfit term is not significant ($p > 0.05$). The response surface model was optimally solved with the same *H* of 600 mm. The optimized parameter combinations obtained were *α=27°, β=22°, V=0.1 m/s, f=6 Hz,* and *A=13 mm*.
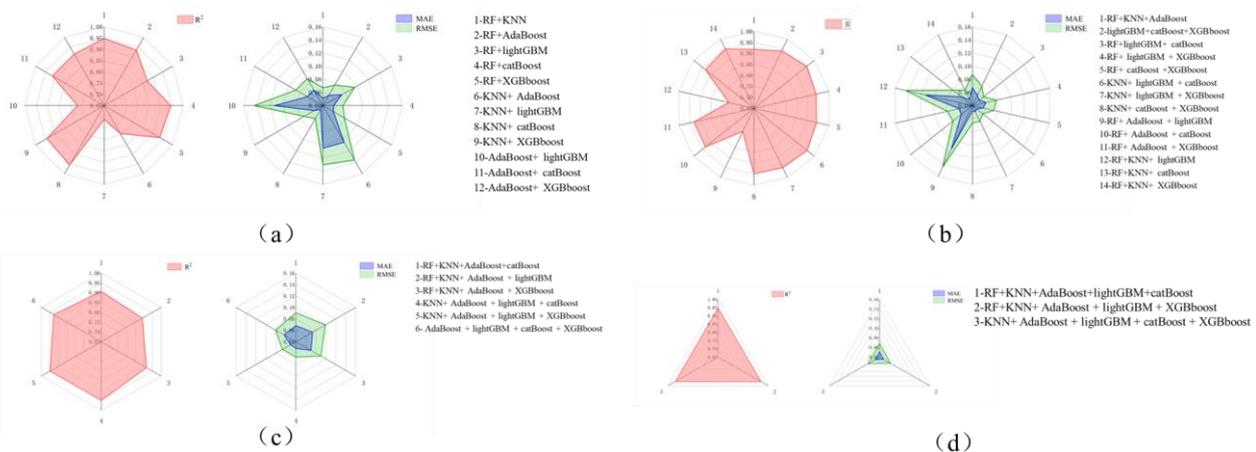
Fig. 9 - Analysis of the prediction accuracy of different combinatorial base learners' graph
(a) two-by-two; (b) three-by-three; (c) four-by-four; (d) five-by-five

Table2

Response Surface Optimization ANOVA

Variation source	Sum of squares	Degree of freedom	Mean square	F value	P value
Model	2.47	27	0.09	8.74	< 0.0001**
α	0.09	1	0.09	9.39	0.005**
β	0.05	1	0.05	5.30	0.0296*
V	0.05	1	0.05	5.33	0.0292*
H	0.30	1	0.30	29.25	< 0.0001**
f	0.95	1	0.95	91.07	< 0.0001**
A	0.08	1	0.08	8.23	0.0081**
αβ	0.06	1	0.06	4.45	0.04*

Variation source	Sum of squares	Degree of freedom	Mean square	F value	P value
αf	0.12	1	0.12	8.96	0.004**
Vf	0.12	1	0.12	8.84	0.004**
VA	0.17	1	0.17	12.53	0.001**
β^2	0.06	1	0.06	4.33	0.04*
V^2	0.02	1	0.02	1.69	0.2
f^2	0.1114	1	0.11	8.00	0.007**
Residual	0.5541	40	0.0139		
Lack of fit	0.5533	35	0.0158	20.81	0.15
Pure error	0.0037	5	0.0007		
Total sum	2.74	53			

Table 3 shows the performance results of the licorice harvester optimized by the three methods. The EL-IGA's optimization outcome considerably lowers resistance and power consumption. In comparison to the RSM, it reduces resistance by 18.16% and specific power consumption by 21.33%; in comparison to the EML-PIGA, it reduces resistance by 11.36% and specific power consumption by 11.19%.

In conclusion, RSM is less useful for solving intricate multi-parameter optimization issues, but the study's EML-IGA can efficiently locate the global optimal solution. The low-order data fitting method, which is prone to falling into local optimal solutions due to its inability to capture the complex relationship between the data, limits the RSM when solving most complex multi-parameter optimization issues. However, EML is good at fitting complex nonlinear relationships, providing an accurate, objective function for the ensuing genetic algorithm optimization, overcoming the limitations of traditional RSM methods, and providing high-precision fitting of design variables and performance metrics. Fast convergence in the optimization process is made possible by combining the IGA method, ensuring that the dominating populations down the population iteration are not killed while maintaining population variety.

Table3

Comparative analysis table of optimization results of different methods

	Resistance /N	Power consumption /KJ
RSM	7411.99	74.4
EML-PIGA	6843.22	65.91
EML-IGA	6065.52	58.53

CONCLUSIONS

This paper uses DEM-MBD coupling simulation, ensemble learning, and an improved genetic algorithm to optimize the licorice harvester's structure and working parameters. It is found that the optimization results of EML-IGA are better than those of traditional RSM, which provides some ideas and methods for multi-parameter optimization.

(1) A coupled DEM-MBD simulation model of the licorice harvester was constructed, which could simulate the actual working condition of the licorice harvester in the soil. The correctness of the simulation model was demonstrated by the DEM model's error of 2.48%, with a relative error of less than 5%, when calibrated using the static soil accumulation angle.

(2) The EML model, trained using a small sample dataset from the simulation model, provides valuable insights into the impact of the quantity and variety of base learners on prediction accuracy. The findings suggest that a diverse range of base learners, in addition to a single model with strong predictive ability, can significantly enhance the predictive capacity of EML.

(3) With a model R^2 of 0.959, an MAE of 0.048, and an RMSE of 0.06, the model has the highest prediction accuracy and accurately represents the mapping relationship between the optimization variables and the optimization metrics when the first layer of the model is KNN + lightGBM + catBoost and the second layer is linear regression.

(4) In comparison to the RSM, EML-IGA reduces resistance by 18.16% and specific power consumption by 21.33%; in comparison to the EML-PIGA, it reduces resistance by 11.36% and specific power consumption by 11.19%. According to the study comparison, EML-IGA is appropriate for solving difficult multi-parameter optimization issues and addresses the shortcomings of RSM.

REFERENCES

[1] Aote, S. S., Pimpalshende, A., Potnurwar, A., & Lohi, S. (2023). Binary Particle Swarm Optimization with an improved genetic algorithm to solve multi-document text summarization problem of Hindi documents. *Engineering Applications of Artificial Intelligence*, *117*. https://doi.org/10.1016/j.engappai.2022.105575

[2] Awuah, E., Zhou, J., Liang, Z., Aikins, K.A., Gbenontin, B.V., Mecha, P., & Makange, N.R. (2022). Parametric analysis and numerical optimisation of Jerusalem artichoke vibrating digging shovel using discrete element method. In *Soil and Tillage Research.* Vol. 219. https://doi.org/10.1016/j.still.2022.105344

[3] Cao, W., Zhang, Z., Fu, Y., Zhao, L., Ren, Y., Nan, T., & Guo, H. (2024). Prediction of arsenic and fluoride in groundwater of the North China Plain using enhanced stacking ensemble learning. *Water Research*, *259* (December 2023). https://doi.org/10.1016/j.watres.2024.121848

[4] de Blas, C. S., Valcarce-Diñeiro, R., Sipols, A. E., Sánchez Martín, N., Arias-Pérez, B., & Santos-Martín, M. T. (2021). Prediction of crop biophysical variables with panel data techniques and radar remote sensing imagery. *Biosystems Engineering*, *205*, 76–92. https://doi.org/10.1016/j.biosystemseng.2021.02.014

[5] Fangping, X., Zhengyang, W., & Xiushan, W. (2020). Calibration of discrete element parameters of soils based on unconfined compressive strength test (基于无侧限抗压强度试验的土壤离散元参数标定) *Transactions of the Chinese Society of Agricultural Engineering,*39-47.

[6] Ge, J., Lei, G., Chen, H., Zhang, B., Chen, L., Bai, M., Su, N., & Yu, Z. (2023). Irrigation district channel dispatch flow prediction based on SHAP importance ranking and machine learning algorithm. *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering*, *39*(13), 113–122. https://doi.org/10.11975/j.issn.1002-6819.202304081

[7] Huang Lvwen, Tian Xu. (2023). Prediction of the soluble solid contents for apple fruit using electrical parameters [J]. Transactions of the Chinese Society of Agricultural Engineering.(基于电学参数的苹果可溶性固形物含量预测) *Transactions of the Chinese Society of Agricultural Engineering*, 252–259.

[8] Junwei, L., Jin, T., Bin, H., Hubiao, W., & Chunyu, M. (2019). Calibration of parameters of interaction between clayey black soil with different moisture content and soil-engaging component in northeast China. *Transactions of the Chinese Society of Agricultural Engineering*, 14–15.

[9] Liao, Z., Yang, Y., Sun, C., Wu, R., Duan, Z., Wang, Y., Li, X., & Xu, J. (2021). Image-based prediction of granular flow behaviors in a wedge-shaped hopper by combing DEM and deep learning methods. *Powder Technology*, *383*, 159–166. https://doi.org/10.1016/j.powtec.2021.01.041

[10] liu Tan, Zhu Hongrui, Y.Q. (2024). Prediction of Photosynthetic Rate of GreenhouseMulti-model Fusion Strategy Tomatoes Based on.(基于多模型融合策略的温室番茄光合速率预测方法) *Transactions of the Chinese Societyfor Agricultural Machinery,*337-345.

[11] Ning Fanghua, Huang Bingqi, Z. X. (2024). Hybrid Flow Shop Batch Scheduling Problem Solving Based on ImprovedGenetic Algorithm. *Software Guide*, *23*(2).

[12] Qiao, Y., Wu, N. Q., He, Y. F., Li, Z. W., & Chen, T. (2022). Adaptive genetic algorithm for two-stage hybrid flow-shop scheduling with sequence-independent setup time and no-interruption requirement. *Expert Systems with Applications*, *208*(July). https://doi.org/10.1016/j.eswa.2022.118068

[13] Ribeiro, M. H. D. M., da Silva, R. G., Moreno, S. R., Mariani, V. C., & Coelho, L. dos S. (2022). Efficient bootstrap stacking ensemble learning model applied to wind power generation forecasting. *International Journal of Electrical Power and Energy Systems*, *136*(October 2021), 107712. https://doi.org/10.1016/j.ijepes.2021.107712

[14] Song Jiannong, W. J., & Wan Lipengcheng, Ll Yonglei, Su Chen. (2021). Simulation of soil tillage characteristics and influence analysis ofparticle sphere type based on EEPA contact model.(基于EEPA接触模型的土壤耕作特性模拟及颗粒球型影响分析) *Journal of China Agricultural University,*193-206.

[15] Wei，LIXinyu, W. K. (2024). Genetic simulated annealing algorithm for resource - constrained human - robot collaborative assembly line balancing problem. *Computer Integrated Manufacturing Systems*. https://doi.org/10.13196/j.cims.2024.0192

[16] Wu, B., Xu, S., Meng, G., & Cui, Y. (2022). Research on structural parameter optimization of elliptical bipolar linear shaped charge based on machine learning. *Heliyon*, *8*(10). https://doi.org/10.1016/j.heliyon.2022.e10992

[17] Yin, F., Dong, H., Zhang, W., Wang, S., Shang, B., & Zhu, Z. (2021). Ability of anaerobic digestion to remove antibiotics contained in swine manure. *Biosystems Engineering*, *212*, 175–184. https://doi.org/10.1016/j.biosystemseng.2021.10.011

[18] Yu, Y., Yan, H., & Ye, C. (2023). Parameter optimization of rectification sill in the forebay of pumping station using BPNN-GA algorithm. (基于BPNN-GA的泵站前池整流底坎参数优化)*Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering*, *39*(14), 106–113. https://doi.org/10.11975/j.issn.1002-6819.202303168

[19] Zhang, R., Han, D., Ji, Q., He, Y., & Li, J. (2017). Calibration Methods of Sandy Soil Parameters in Simulation of Discrete Element Methodn（离散元模拟中沙土参数标定方法研究）*Nongye Jixie Xuebao/Transactions of the Chinese Society for Agricultural Machinery*, *48*(3), 49–56. https://doi.org/10.6041/j.issn.1000-1298.2017.03.006

[20] Zhang, X., Kang, M., Shi, Z., Yan, J., Liu, X., Guo, L., & Wang, M. (2024). Design and experiments of a double drum surface residual film recycling machine after harvest. *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering*, *40*(14), 1–13. https://doi.org/10.11975/j.issn.1002-6819.202401037

[21] Zongquan, L. Z. (1997). Prediction Method for Reinforced Concrete Corrosion-induced Crack Based on Stacking Integrated Model Fusion. *Journal of Chinese Society for Corrosion and Protection*.