A Hybrid SETO-GBDT Model for Efficient Information Literacy System Evaluation

Jiali Dai[1], Hanifah Jambari[2], Mohd Hizwan Mohd Hisham[3]
College Office, Wenzhou Polytechnic, Wenzhou 325000, Zhejiang, China[1]
Faculty of Social Sciences and Humanities, Universiti Teknologi Malaysia, Johor, 81310, Malaysia[1, 2, 3]

***Abstract*—Information literacy (IL) is essential for vocational education talents to thrive in the modern information age. Traditional assessment methods often lack quantitative precision and systematic evaluation models, making it difficult to accurately measure IL levels. This paper aims to develop a robust, data-driven model to assess information literacy in vocational education talents. The goal is to improve the accuracy and efficiency of IL evaluations by combining machine learning techniques with optimization algorithms. The proposed method integrates the Stock Exchange Trading Optimization (SETO) algorithm with the Gradient Boosting Decision Tree (GBDT) to construct the SETO-GBDT model. This model optimizes parameters such as the number of decision trees and tree depth. A comprehensive evaluation index system for IL is built, focusing on learning attitude, process, effect, and practice. The SETO-GBDT model was trained and tested using real-world data on IL indicators. The SETO-GBDT model outperformed traditional models such as Decision Tree, Random Forest, and GBDT optimized by other algorithms like SCA and SELO. Specifically, it achieved an RMSE of 0.13, an R² of 0.98, and reduced evaluation time to 0.092 s, demonstrating superior accuracy and efficiency. The research concludes that the SETO-GBDT model offers a significant improvement in evaluating IL for vocational education talents. The model's high accuracy and reduced evaluation time make it an effective tool for assessing and enhancing information literacy, aligning with the educational goals of developing well-rounded, information-savvy professionals.**

***Keywords*—Vocational education; talent; information literacy; system building; educational evaluation; gradient augmentation; decision tree**

I. INTRODUCTION

China has implemented various programs to enforce the systematic advancement of educational evaluation reform. The primary focus of the current curriculum reform is to develop scientific core literacy. The goal is to transform the curriculum to enhance students' comprehensive ability to use information technology to solve problems. This will enable students to become well-rounded individuals with high-quality technical skills and moral, intellectual, physical, social and aesthetic development [1]. Educational talent assessment plays a crucial role in talent education. It involves designing an evaluation index system and educational activities to determine the reasonableness of the educational process, appropriateness of the educational methods used, and whether the expected educational outcomes are achieved [2]. In vocational education, the rapid advancement of information technology has led to the constant flow of information. In this vast and complex information landscape, the ability to distinguish, summarize, and synthesize information has become an increasingly important challenge [3]. For college students, the skill of efficiently and accurately finding the information they need within a limited time frame has become an essential foundational skill. The enhancement of information literacy among vocational education talents involves developing their vocational information literacy skills, including vocational skills, information skills, comprehensive skills, and other vocational information literacy skills. This enables them to acquire the necessary knowledge and adapt to the rapid development of the information society [4].

The evaluation of vocational education talent information literacy is a crucial tool to enhance the information literacy of vocational education talent. Employing appropriate assessment methods may facilitate students' growth and ensure that their information literacy meets the intended objectives [5]. The present research on talent information literacy focuses mostly on three areas: defining information literacy, constructing talent information literacy systems, and assessing talent information literacy [6]. Fei and Erjun [7] studied the eight aspects of information literacy, i.e., skillful use of information tools, correct access to information, proper handling of information, timely generation of information, good at creating information, maximizing the benefits of information, strengthening information collaboration, enhancing information immunity, etc.; Esfandiari and Arefian [8] investigated four information literacy evaluation indexes for undergraduate-type students, i.e., information awareness, information competence, information evaluation, and information ethics; Riithi and Kimani [9] analyzed the academic attention to information literacy education and found that 2012 to 2014 was a period of rapid growth in information literacy research; Ganesan and Gunasekaran [10] elaborated on the content of information literacy education in applied schools and gave strategies and methods for cultivating students' information literacy; Vianna and Caregnato [11] used hierarchical analysis methods to construct a system of information literacy for talents and studied the corresponding cultivation methods; Faber [12] used a simple decision tree to fit the nonlinear mapping relationship between talent information literacy indicators and assessment values. The examination of the present research literature on information literacy reveals the existence of the following issues: a) The present assessment of talent information literacy is still in its early developmental stage, focusing mostly on qualitative analysis of the meaning of information literacy and the importance of the study, while lacking quantitative analysis [13]; b) The present assessment methods are not suitable for constructing the information literacy assessment model due to the large dimensionality of the information literacy evaluation index system. The selection of the talent information literacy evaluation index is insufficiently thorough and lacks a systematic approach [14].

The advancement of integrated learning technology has led to the adoption of efficient integrated decision trees to improve assessment models. This has become a prominent area of future development and research. However, the performance of the integrated learning algorithm is limited by the parameter settings. Therefore, optimizing the algorithm through hyper-parameter search has emerged as a method to enhance the assessment algorithm [15]. This paper presents a method for assessing the information literacy of vocational education talents. The method combines literature analysis and the development of intelligent algorithms, specifically using an intelligent optimization algorithm-integrated decision tree framework. This paper presents a talent information literacy assessment model for vocational education talents by analyzing the problem of constructing and evaluating an information literacy system. The model is designed using a framework that combines the gradient enhancement decision tree and the stock market trading optimization algorithm. It is then applied to the development of information literacy in vocational education talents. By doing rigorous experimental research, this study confirms that the proposed strategy is indeed superior.

II. Information Literacy System

A. Analysis of the Process

The purpose of developing an information literacy system for vocational education skills is to enhance students' overall abilities in the field of information technology. This includes fostering their awareness of information, understanding of information, competence in handling information, and adherence to ethical standards related to information (Fig. 1). Developing an information literacy system for vocational education can enhance students' capacity to adapt to the demands of work and life in the digital age, while also improving their vocational competitiveness and lifelong learning skills [16].

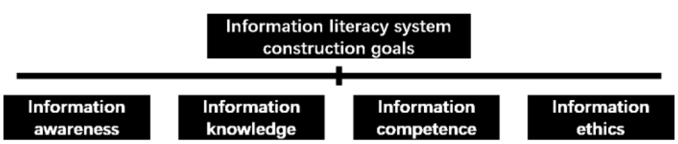

Fig. 1. Objectives of information literacy system construction for vocational education talents.

The process of constructing a vocational education talent literacy system involves various components such as curriculum design, teaching methods, teaching resources, teacher training, assessment, and feedback [17], as depicted in Fig. 2. This process is driven by the goal and importance of developing vocational education talent literacy.

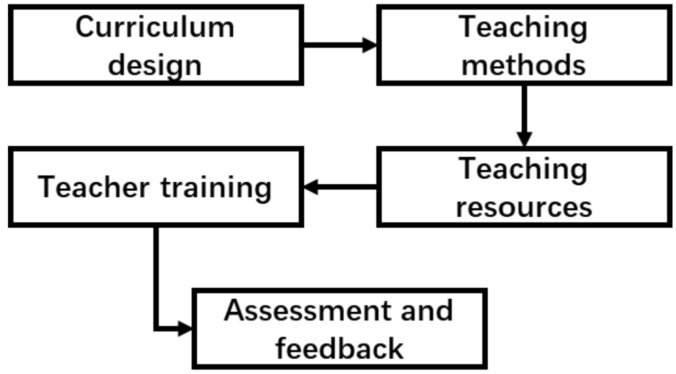

Fig. 2. Construction process of information literacy system for vocational education talents.

B. Constructive Thinking Refers to the Process

The construction process of the vocational education talent information literacy system involves extracting talent literacy assessment indexes from four aspects: learning attitude, learning process, learning effect, and final practice. These indexes serve as the first-level indexes. Through the refinement of the evaluation process, second-level indexes are extracted to construct the vocational education talent literacy system. Fig. 3 illustrates the concept of this construction process.

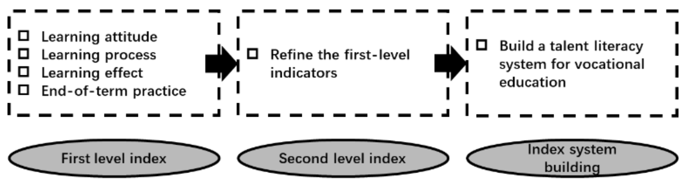

Fig. 3. Constructing information literacy system for vocational education talents.

C. Construction of Talent Information Literacy System

This paper adheres to the principles of objectivity, comprehensiveness, focus, and practicality in constructing an information literacy system for vocational education talents (Fig. 4). It selects relevant indicators from four aspects: information literacy learning attitude, learning process, learning effect, and final practice, as illustrated in Fig. 4. The indicators for learning attitude encompass class attendance and the number of assignments submitted. The indicators for learning process involve the organization of information literacy materials, the design of information literacy papers, information retrieval, information management, and display presentation production. The indicators for learning effects consist of teacher information literacy training and student information feedback. The final practice indicators encompass information literacy application and examination results [14].

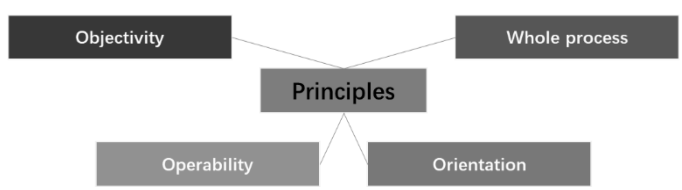

Fig. 4. Principles for the selection of evaluation system indicators.

Diagram depicting the creation of the evaluation system, as seen in Table I.

TABLE I. EVALUATION OF SYSTEM CONSTRUCTION

No.	First level	Var.	Second level	Var.
1	Learning attitude	A	Class attendance	A1
			Number of assignments turned in	A2
2	Learning process	B	Material arrangement	B1
			Information literacy paper design	B2
			Information retrieval	B3
			Information management	B4
			Display and presentation production	B5
3	Learning effect	C	Literacy training for teachers	C1
			Feedback on students' information	C2
4	End-of-term practice	D	Information literacy application	D1
			Test scores	D2

III. EVALUATING THE INFORMATION LITERACY SKILLS

A. Talent Information Literacy Assessment Framework

The method for assessing the information literacy of vocational education talents involves using the information literacy assessment index as input for the evaluation model. The output is a comprehensive assessment score. The assessment model is constructed using the integrated learning method based on the gradient boosting decision tree algorithm. The hyperparameters of the gradient boosting decision tree are optimized using a stock market trading optimization algorithm. The specific framework is illustrated in Fig. 5.

The talent information literacy evaluation model construction framework analysis reveals that the research on constructing and assessing the talent information literacy system in vocational education is a crucial technology. This research utilizes an enhanced integrated learning algorithm to establish a mapping relationship between the assessment index value of information literacy and the assessment scores, as depicted in Fig. 6. This research utilizes gradient boosting decision tree to establish the mapping connection and enhances the assessment accuracy of the talent information literacy evaluation model by optimizing the hyperparameters of the GBDT technique using a stock market trading optimization algorithm.

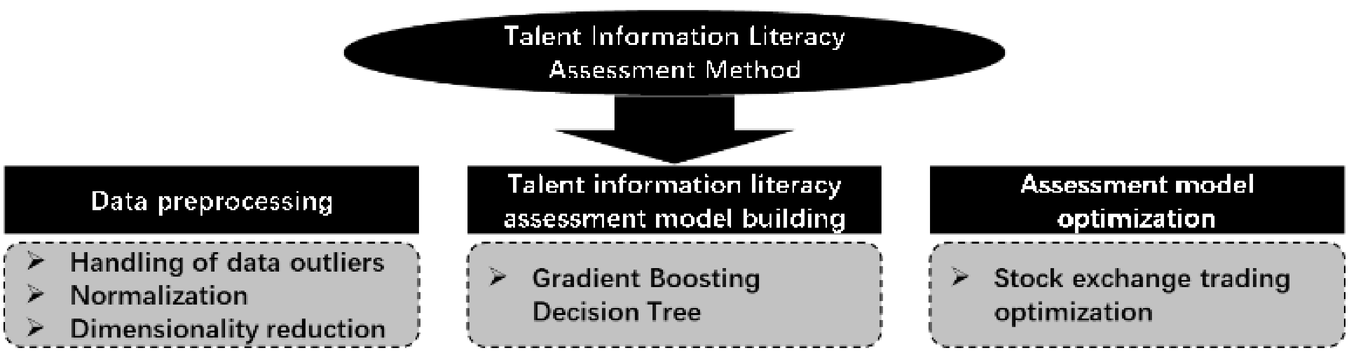

Fig. 5. Talent information literacy evaluation model construction framework.

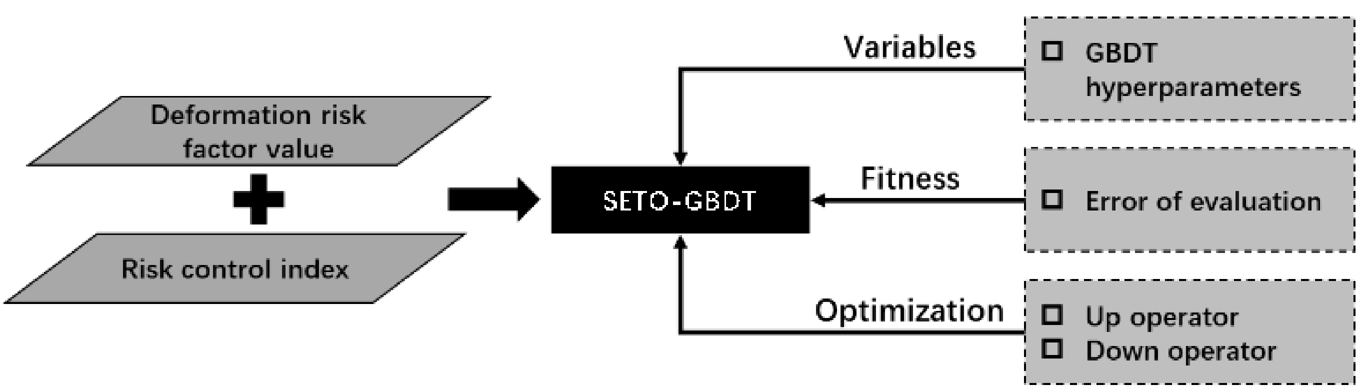

Fig. 6. Information literacy assessment method for vocational education talents based on the improved integrated learning model.

B. Gradient Boosting Decision Tree

1) The theory of gradient enhancing decision trees: The Gradient Boosting Decision Tree (GBDT) [18] is a popular machine learning approach that is commonly employed for solving regression and classification issues. It enhances the precision of predictions by repeatedly creating several decision trees and decreasing the loss function via gradient descent. The primary objective of each new decision tree is to rectify the residuals of the preceding tree, which refers to the disparity between the actual value and the current model's prediction. The fundamental concept behind Gradient Boosting Decision Trees (GBDT) is to construct a robust prediction model by amalgamating numerous feeble learners, typically decision trees. The structure of this model is illustrated in Fig. 7.

In this paper, the GBDT algorithm is chosen to build an information literacy assessment model for vocational education talents in the form of regression tree. The specific form of decision tree is as follows:

$$T\left(x;c,R\right)=\sum_{v=1}^{M'}c_v I\left(x\in Rv\right) \tag{1}$$

$$I=\begin{cases}0 & x\notin R_v\\ n & x\in R_v\end{cases} \tag{2}$$

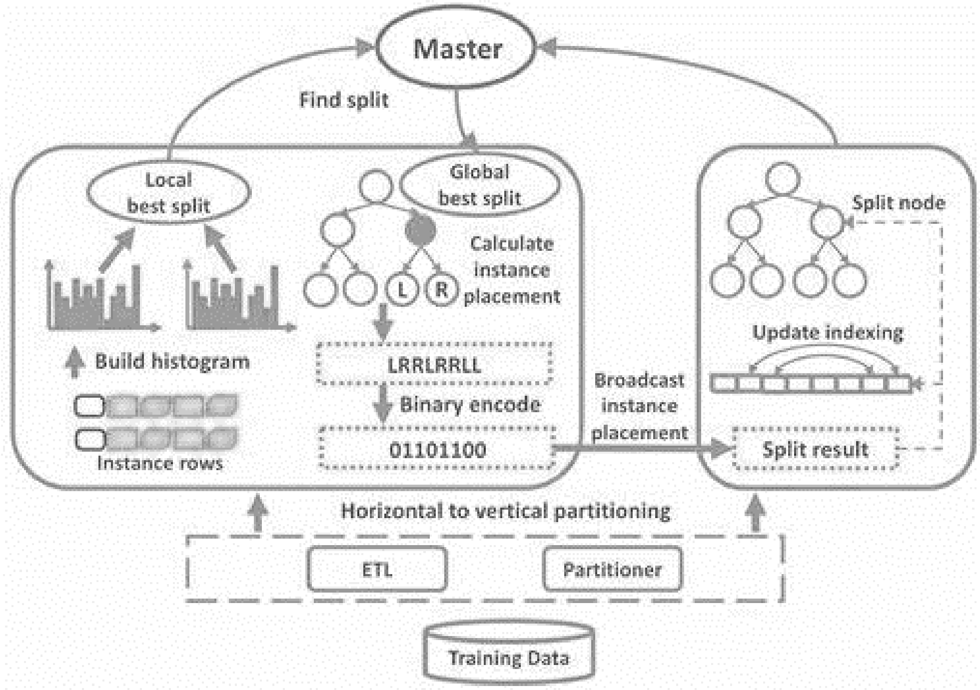

Fig. 7. GBDT structure.

Where, $R=\left\{R_1,R_2,\cdots,R_M^t\right\}$ is the decision tree leaf; M^t is the number of leaf nodes; I is the difference function; $c=\left\{c_1,c_2,\cdots,c_{Mt}\right\}$, $c_v=mean\left(y\middle|x\in R_v\right)$ represent the output characteristic mean of the samples in the leaf space. The GBDT model is a combination of multiple decision trees with the following structure:

$$F\left(x\right)=\sum_{ix=1}^{N'}T_{ix}\left(x;c,R\right) \tag{3}$$

where T_{ix} represents the ixth tree and $ix=1,2,\cdots,N^t$.

During each iteration, a decision tree T_m is added to the decision model based on the previous iteration number in the following form:

$$T_m=\arg\min_T\sum_{k=1}^{n^t}L\left(y_k,\sum_{k=1}^{m-1}T_j\left(x_k\right)+T\left(x_k\right)\right) \tag{4}$$

Where $L\left(\cdot\right)$ denotes the loss function, j is the number of iterations, k is the number of samples in the training set, x_k and y_k denote the training samples. The loss function is set to the least squares function, and the m^{th} tree is built on the residuals of the sum of the decision trees in the previous iteration, and the GBDT model is constructed as follows after m iterations:

$$F_m\left(x\right)=F_{m-1}\left(x\right)+\arg\min_T\sum_{k=1}^{n^t}L\left(y_k,F_{m-1}\left(x_k\right)+T\left(x_k\right)\right) \tag{5}$$

The minima of the GBDT model are calculated by the gradient descent method and the negative gradient direction of the loss function at the current F_{m-1} is set to be the direction of the maximum descent gradient:

$$F_m\left(x\right)=F_{m-1}\left(x\right)+r_m\sum_{k=1}^{n^t}\nabla F_{m-1}L\left(y_k,F_{m-1}\left(x_k\right)\right) \tag{6}$$

$$r_m = \arg\min_r \sum_{k=1}^{n^t} L\left(y_k, F_{m-1}(x_k) - r\frac{\partial L(y_k, F_{m-1}(x_k))}{\partial F_{m-1}(x_k)}\right) \quad (7)$$

In order to avoid the fitting phenomenon of GBDT model, the model learning rate was used to determine the model:

$$F_m(x) = F_{m-1}(x) + v r_m T_m(x) \quad (8)$$

where v is the learning rate of the GBDT model.

2) GBDT process steps: The GBDT principle involves a sequential process outlined in a flowchart (Table II). The steps are as follows: 1) initialize the model, 2) calculate the residuals, 3) build the decision tree, 4) update the model, and 5) output the model until a predetermined number of iterations is reached or the stopping condition is met [19].

TABLE II. GBDT FLOW CHART

Algorithm 1: GBDT Model
1. Initialize model.
2. Calculate residuals.
3. Construct a decision tree.
4. Update the model.

3) GBDT Advantage: The Gradient Boosting Decision Tree (GBDT) has several advantages: 1) It exhibits a high level of precision in its prediction capacity; 2) It is capable of directly handling various sorts of data; 3) It demonstrates resilience and generalization ability; 4) It efficiently handles large-scale data; and 5) It can effectively train unbalanced data, as seen in the accompanying Fig. 8.

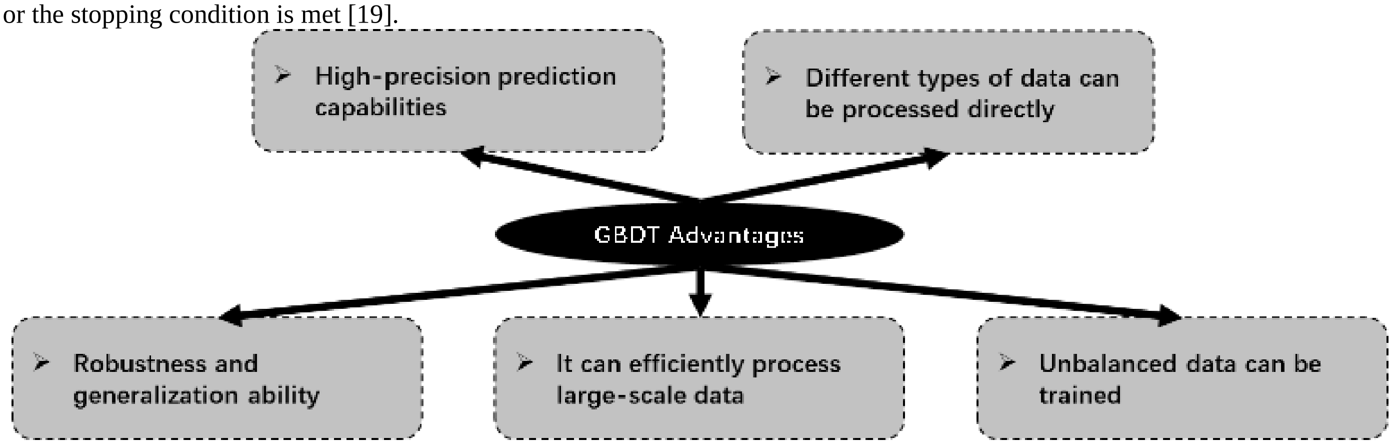

Fig. 8. GBDT advantages

4) Applications of gradient boosting decision trees: Gradient Boosting Decision Trees (GBDT) demonstrates exceptional performance in several practical applications (Fig. 9), such as financial risk management, stock market forecasting, medical diagnostics, and natural language processing [20]. Due to its strength and versatility, this tool is highly used by data scientists and machine learning engineers for addressing intricate prediction issues (Fig. 10) [19].

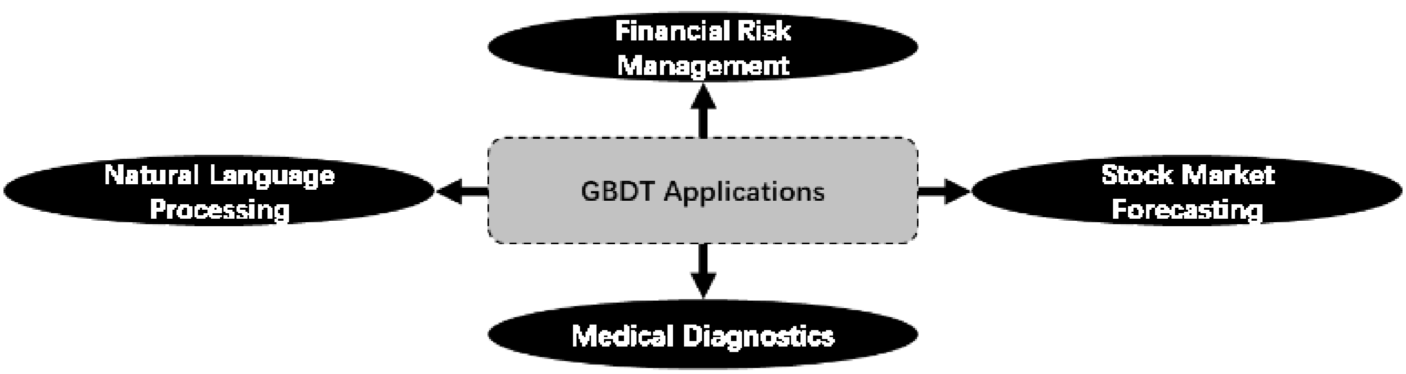

Fig. 9. GBDT application.

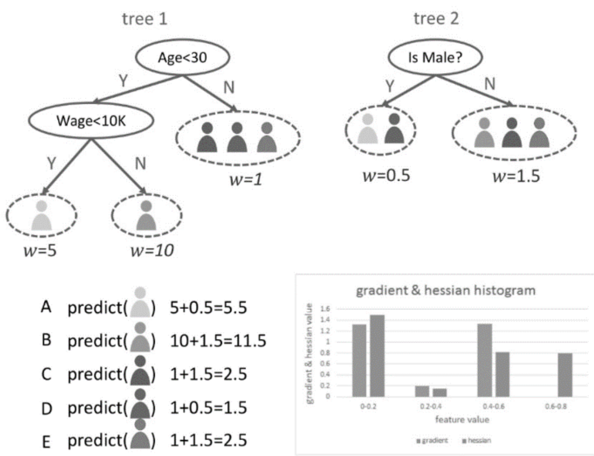

Fig. 10. GBDT problem solving approach.

C. The SETO Optimized Gradient Boosting Decision Tree (GBDT) Model

1) SETO algorithm: A swarm intelligence optimization algorithm known as Stock Exchange Trading Optimization (SETO) [21] takes its cues from the ever-changing stock market and its trading patterns to determine which stock is most likely to maximize profit. Here, each stock is seen as a possible solution to the problem. The method repeatedly optimizes operators such as rise, fall, and exchange operations. Ultimately, the share that yields the highest profit is identified as the ideal answer.

a) Initialization of the population

$$s_{ij} = l_{ij} + \phi_{ij} \cdot \left(u_{ij} - l_{ij} \right) \tag{9}$$

ϕ_{ij} represents a random integer in the interval [0, 1], u_{ij} and l_{ij} signify the upper and lower bounds of the search space, respectively, and s_{ij} symbolizes the j^{th} dimension associated with the i^{th} person.

The following is the equation for calculating the earnings value of a stock, which is used to analyze individual stocks:

$$f_i = f\left(S_i\right) = f\left\{s_{i1}, s_{i2}, \cdots, s_{iD}\right\} \tag{10}$$

where f_i denotes the fitness value of the i^{th} stock individual S_i .

There are specific numbers of sellers and buyers for every stock in the stock market. The first traders are defined using a random initialization process. Here nf_i is how to compute the normalized fitness value in order to accomplish this mechanism:

$$nf_i = \frac{f_i - \min\left(M\right)}{\sum_{k=1}^{N}\left(f_k - \min\left(M\right)\right)}, M = \left\{ f_k \middle| k = 1, 2, \cdots, N \right\} \tag{11}$$

S_i The number of traders is calculated as follows:

$$T_i = \left[nf_i \times T \right] \tag{12}$$

T represents the total number of traders, whereas T_i denotes the trading volume of stock S_i. Here S_i is how to figure out how many people are buying and selling stocks:

$$b_i = \left[r \times T_i \right] \tag{13}$$

$$s_i = T_i - b_i \tag{14}$$

b_i and s_i represent the quantities of buyers and sellers, respectively, whereas r is a random variable uniformly distributed between $[0,1]$.

b) Ascent operation operator: The upward action mostly emulates the appreciation of the stock price. At this juncture, the stock may ascend to a greater valuation, and the peak price can attain the ideal threshold. The equation for replicating the upward action operator is articulated as follows:

$$S_i\left(t+1\right) = S_i\left(t\right) + R \times \left(S^g\left(t\right) - S_i\left(t\right) \right) \tag{15}$$

$S_i\left(t\right)$ represents the i^{th} stock person for the t^{th} iteration, $S^g\left(t\right)$ signifies a D-dimensional random vector, and $r_j \in R$ indicates the ideal solution for the t^{th} iteration.

The parameter R enhances the degree of random variation to assist individuals in evading local optima and exploring broader geographical areas, while $r_j \in R$ is specified as follows:

$$r_j = U\left(0, pc_i \times d_1\right) \tag{16}$$

U produces evenly dispersed random numbers within the range of $\left[0, pc_i \times d_1\right]$. pc_i is the bid-ask ratio of the stock S_i, and d_1 denotes the normalized distance between $S_i\left(t\right)$ and $S^g\left(t\right)$:

$$d_1 = \frac{\sqrt{\sum_{j=1}^{D}\left(S_j^g\left(t\right) - S_{ij}\left(t\right)\right)^2}}{ub - lb} \tag{17}$$

ub and lb represent the upper and lower limits of the search space, respectively. Typically, an increase in demand for a stock correlates with an appreciation in its value. The parameter pc_i mimics the effect of stock growth demand and delineates stock demand based on the overall number of purchasers, calculated as follows:

$$pc_i = \frac{b_i}{s_i + 1} \tag{18}$$

In order to avoid pc_i crossing the boundary, the parameter pc_i is limited to the range $[0, 2]$, which is calculated as follows:

$$pc_i = \min\left(\frac{b_i}{s_i + 1}, 2\right) \tag{19}$$

In the ascending phase, stock demand escalates, resulting in a rise in buyers and a reduction in sellers:

$$b_i = b_i + 1 \tag{20}$$

$$s_i = s_i - 1 \tag{21}$$

c) Falling operation operator: Most of the time, the decline operation operator will mimic a falling stock price using the following equation:

$$S_i(t+1) = S_i(t) - W \times \left(S_i^l(t) - S_i(t)\right) \tag{22}$$

$S_i^l(t)$ represents the current local optimal solution of the ith stock, W signifies a D-dimensional random vector, and $w_j \in W$ is defined as follows:

$$w_j = U\left(0, nc_i \times d_2\right) \tag{23}$$

Where U generates uniformly distributed random numbers in the range $[0, nc_i \times d_2]$. nc_i is the sell-buy ratio of the stock S_i, and d_2 denotes the normalized distance between $S_i(t)$ and $S_i^l(t)$:

$$d_2 = \frac{\sqrt{\sum_{j=1}^{D}\left(S_{ij}^l(t) - S_{ij}(t)\right)^2}}{ub - lb} \tag{24}$$

$$nc_i = \min\left(\frac{s_i}{b_i + 1}, 2\right) \tag{25}$$

In the decline phase, the supply of stocks escalates. In each repetition, the number of vendors rises while the number of customers diminishes during the decline phase:

$$s_i = s_i + 1 \tag{26}$$

$$b_i = b_i - 1 \tag{27}$$

d) Exchange operation operator: During the trading phase, the trader employs the most lucrative stock to substitute the least expensive stock. During this phase, the trader divests from the least performing stock and acquires the highest performing stock. This operational method enables traders to attract stocks. The least favorable stock is obtained as follows:

$$\begin{aligned} &S_{worst} = S_w \; where \quad f\left(S_w\right) < f\left(S_j\right) \\ &\forall j = 1, 2, \cdots, N, w \neq j \end{aligned} \tag{28}$$

Subsequently, the least favorable stock queue eliminates one seller and incorporates it into the most favorable stock queue and the ideal stock. S_{best} definition is derived as follows:

$$\begin{aligned} &S_{best} = S_b \; where \quad f\left(S_b\right) < f\left(S_j\right) \\ &\forall j = 1, 2, \cdots, N, b \neq j \end{aligned} \tag{29}$$

The exchange operation operator augments the population size. The operation reduces the quantity of suppliers while augmenting the quantity of purchasers. Consequently, the buyer-seller ratio escalates, thereby enhancing the probability of the stock appreciating.

e) RSI calculation: We use the RSI indicator to recognize when a stock is rising or falling. As the RSI value increases, SETO behaves up or down modeled as follows:

$$\begin{cases} \text{rising} & RSI \leq 30 \\ \text{falling} & RSI \geq 70 \\ p \times \text{rising} + (1-p) \times \text{falling} & 30 < RSI < 70 \end{cases} \tag{30}$$

When p represents a binary random variable, and $p \in \{0,1\}$ is calculated as follows:

$$p = \begin{cases} 1 & rand \geq 0.5 \\ 0 & else \end{cases} \tag{31}$$

$rand$ represents a random number inside the interval [0, 1]. The RSI for the ith stock is computed as follows:

$$RSI = 100 - \frac{100}{1 + RS} \tag{32}$$

Relative intensities were computed using the simple moving average method:

$$RS = \frac{\sum_{i=1}^{K} P_i}{\sum_{i=1}^{K} N_i} \tag{33}$$

P_i and N_i represent upward and downward price fluctuations, respectively. K represents the RSI trading time frame. The equations for P_i and N_i are as follows:

$$P_i = \begin{cases} 1 & if\left(f_i(t) - f_i(t-1) \right) > 0 \\ 0 & otherwise \end{cases} \tag{34}$$

$$N_i = \begin{cases} 1 & if\left(f_i(t-1) - f_i(t) \right) > 0 \\ 0 & otherwise \end{cases} \tag{35}$$

where, $f_i(t)$ and $f_i(t-1)$ denote the fitness values for the current versus previous iteration counts, respectively.

f) Algorithmic steps: The SETO algorithm's location updating approach is shown in the pseudo-code included in Table III.

TABLE III. SETO ALGORITHM PSEUDO-CODE

Algorithm 2: SETO Algorithm
1. Initialize AOA parameters;
2. Initialize population of shares;
3. Evaluate initial population and update best share with best value;
4. While t <= tmax do
5. For each share do
6. If RSI <= 30
7. Carry out rising operator;
8. Elseif RSI >= 70
9. Carry out falling operator;
10. Else
11. Carry out rising and falling phase;
12. End
13. Carry out exchange phase;
14. Calculate RSI;
15. End
16. Evaluate object and update best object;
17. t = t + 1;
18. End
19. Output best solution.

2) SETO-GBDT: This paper utilizes the SETO algorithm to optimize the parameters of the GBDT model [22]. These parameters include the number of decision trees (Para1), the maximum depth of the tree (Para2), the minimum number of samples required for internal nodes (Para3), the minimum number of samples required for leaf nodes (Para4), and the optimal number of segmented features (Para5). The optimization process aims to minimize the regression error of the GBDT model, as demonstrated in Table IV.

D. *Application of SETO-GBDT Model in the Assessment of Information Literacy of Vocational Education Talents*

This research applies the SETO algorithm optimization GBDT model to design a vocational education talent information literacy assessment model, aiming to tackle the problem of vocational education talent information literacy assessment. The information literacy assessment method for vocational education talents, based on the SETO-GBDT model, consists of two main components: the development of an information literacy assessment index system for vocational education talents, and the creation of an information literacy assessment model for vocational education talents. The specific steps for implementing this method are illustrated in Fig. 11.

Fig. 11 illustrates the process of constructing an information literacy assessment index system for vocational education talents. The first part involves analyzing the development of information literacy in vocational education talents and using this analysis to design the information literacy assessment index system. The second part focuses on standardizing the data of the information literacy assessment index for vocational education talents, with the index value serving as input and the information literacy assessment value as output. The literacy assessment indicator data is standardized using the SETO algorithm to optimize the GBDT parameters. This algorithm is used to train the mapping relationship between the indicator value and the assessment value of vocational education talents' information literacy assessment.

TABLE IV. PSEUDO-CODE OF SETO-GBDT ALGORITHM

Algorithm 3: GBDT based on SETO algorithm
1. Determine optimized variables, including Para 1–5;
2. Set SETO algorithm parameters;
3. Initialize stock population;
4. Calculate fitness of stock using Error, and update best stock;
5. While t <= tmax
6. Calculate RSI value;
7. If RSI <= 30
8. Carry out rising operator;
9. Else if RSI >= 70
10. Carry out falling operator;
11. Else Carry out rising and falling phase;
12. End
13. Carry out exchange phase;
14. End
15. Output best parameters of GBDT model;
16. Build SETO-GBDT model.

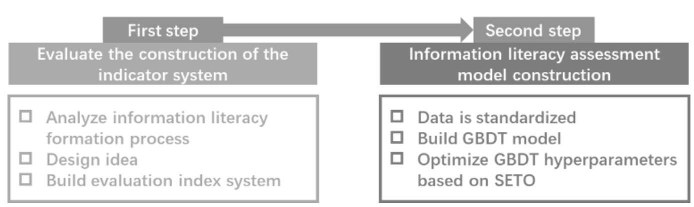

Fig. 11. Step-by-step diagram of the information literacy assessment model for vocational education talents combined with SETO-GBDT.

IV. Data Analysis

This paper aims to assess the effectiveness of the SETO-GBDT model in evaluating the information literacy of vocational education talents. We utilize a dataset consisting of information literacy assessment indexes of vocational education talents and compare and analyze the performance of the SETO-GBDT model with the GBDT model optimized by the SCA [23], SELO [24], HBO [25], and LFD [26] algorithms.

A. Environment, Data, and Algorithm Settings

The SETO-GBDT model is used to assess the information literacy of vocational education talents through a simulation experiment in the Windows 10 environment. The visualization software used is Matlab 2022a, the method programming software is Python 3.8, and the fundamental algorithm is implemented in C++.

The data set of indicators for assessing information literacy in vocational education was gathered by methods such as literature data analysis, case study analysis, comparison analysis, and questionnaire survey (Fig. 12). The data from the study subjects were randomly split into three groups: 70% for training, 15% for testing, and 15% for validation. The model's average evaluation indexes were then calculated using the ten-fold cross-validation method.

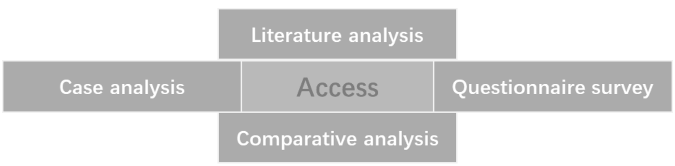

Fig. 12. Data access.

The information literacy evaluation algorithm for vocational education abilities based on the SETO-GBDT model utilizes several comparison algorithms, such as decision tree, RF, AdaBoost, GBDT, and GBDT algorithms optimized by SCA, SELO, HBO, and LFD. The particular parameter values may be found in Table V and Table VI. The algorithms SCA, SELO, HBO, and LFD each have 100 populations and a maximum iteration number of 1000.

TABLE V. Parameter Settings of the Contrast Evaluation Algorithm

No.	Algorithms	Parameter settings
1	Decision tree	Maximum number of splits is 4
2	Random forest	N_tree=500, m_try=floor(80.5)
3	AdaBoost	Regressors number is 15, Iteration is 50
4	GBDT	Decision tree number is 48, maximum depth of tree is 10, samples for internal nodes is 18, samples required for leaf nodes is 1, optimal segmentation features is 9.

TABLE VI. Comparison Optimization Algorithm Parameter Settings

No.	Algorithms	Parameter settings
1	SCA	a=2, r1=1-2t/G, r2=[0,2π], r3=[0,2], r4=[0,1]
2	SELO	P=2, O=3, rp=0.999, rk=0.1, prob=0.999
3	HBO	C=G/25, p1=1-t/G, p2=p1+(1-p1)/2
4	LFD	Threshold=2, CSV=0.5, β=1.5
5	SETO	T=100

B. GBDT Model Optimization Results

The average index value of each algorithm is statistically obtained through the ten-fold cross-validation method, and the specific results are shown in Table VII, Table VIII and Fig. 13.

TABLE VII. Optimization Results of different Optimization Algorithms to Optimize the GBDT Model

No.	Algorithms	Opti. value	Opti. time	Iter. num
1	SCA-GBDT	2.668	3.72	1000
2	SELO-GBDT	1.460	3.33	968
3	HBO-GBDT	0.510	3.45	755
4	LFD-GBDT	0.367	3.10	631
5	SETO-GBDT	0.125	2.71	400

TABLE VIII. Results of Different Optimization Algorithms to Optimize the Parameters of GBDT Model

No.	Algorithms	Para1	Para2	Para3	Para4	Para5
1	SCA-GBDT	30	10	10	2	12
2	SELO-GBDT	41	9	21	3	7
3	HBO-GBDT	49	10	10	3	12
4	LFD-GBDT	44	7	19	5	10
5	SETO-GBDT	50	10	25	5	8

Table VIII presents a comparison of the optimization accuracy, optimization time, and convergence number results of different optimization algorithms for optimizing the GBDT model. From Table VIII, it can be seen that in terms of optimization accuracy, the GBDT evaluation model based on SETO algorithm has the highest accuracy of 0.125, followed by LFD-GBDT, HBO-GBDT, SELO-GBDT, and SCA-GBDT models; in terms of optimization time, the SETO-GBDT model has the lowest evaluation time of 2.71 s; in terms of the number of times of convergence, the SETO-GBDT model converged to the optimal value in 400 times.

The results of optimizing GBDT model parameters with different optimization algorithms are given in Fig. 13. The results of optimizing GBDT model parameters based on SETO algorithm: number of decision trees Para1=50, maximum depth of tree Para2=10, minimum number of samples required for internal nodes Para3=25, minimum number of samples required for leaf nodes Para4=5, and optimum number of segmentation features Para5=8.

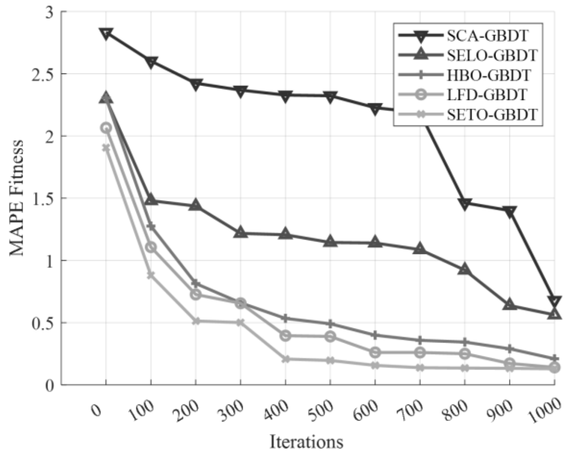

Fig. 13. Convergence curve of GBDT model optimized by different optimization algorithms.

Fig. 13 gives the convergence curve of GBDT model optimized by different optimization algorithms. In Fig. 13, it can be seen that the convergence curve of the optimized GBDT model based on SETO algorithm converges to 0.125 at the 400th iteration.

C. Evaluation of Model Comparison Results

In order to avoid unexpected results of the experiment, 10 independent tests were conducted and the RMSE, R2, training time, and evaluation time averages of the decision tree, RF, AdaBoost, GBDT, and SETO-GBDT algorithms were counted, as shown in Table IX.

TABLE IX. STATISTICS AND COMPARISON OF THE RESULTS OF THE ASSESSMENT INDICATORS FOR THE CONTRASTING ASSESSMENT ALGORITHMS

No.	Evaluation models	RMSE	R^2	Training T/s	Evaluation T/s
1	Decision tree	1.37	0.76	2.30	0.189
2	RF	0.88	0.86	3.94	0.166
3	AdaBoost	0.36	0.93	3.25	0.132
4	GBDT	0.31	0.96	4.45	0.104
5	SETO-GBDT	0.13	0.98	2.71	0.092

Table IX presents the statistics and comparisons of the evaluation index results for the decision tree, RF, AdaBoost, GBDT, and SETO-GBDT algorithms. The SETO-GBDT algorithm performs the best in terms of RMSE, achieving a value of 0.13. It also outperforms the other algorithms in terms of R2, with a value of 0.98. The decision tree algorithm has the shortest training time, taking only 2.30 seconds. On the other hand, the SETO-GBDT algorithm has the shortest evaluation time, which is 0.092 seconds.

V. CONCLUSION

This paper addresses the issue of assessing information literacy in vocational education. It proposes a model for assessing information literacy in vocational education based on SETO-GBDT. The model is validated through literature analysis, case study analysis, comparative analysis, and questionnaire survey. The findings are as follows:

- The SETO method enhances the convergence accuracy and decreases the optimization time of the GBDT model compared to other optimization techniques, while also accelerating the convergence process.
- The SETO-GBDT model outperforms other evaluation models in terms of evaluation error RMSE, R2 value, and evaluation time. The RMSE is 0.13, the R2 value is 0.98, and the evaluation time is 0.092s.
- The experimental findings confirm the accuracy of the SETO-GBDT model in evaluating the impact.

The superior performance of the SETO-GBDT model is evident in its higher convergence accuracy and shorter optimization time, making it a valuable tool for educational institutions seeking to assess and enhance the information literacy of their students. This model not only streamlines the evaluation process but also aligns with the goal of fostering well-rounded, information-savvy professionals in today's fast-paced, data-driven society.

REFERENCES

[1] Munavalli S B .Impact of information literacy skill assessment to explore learning resources[J].Journal of Library and Information Communication Technology, 2023.

[2] Juan Bartolomé, Garaizar P .Design and Validation of a Novel Tool to Assess Citizens' Netiquette and Information and Data Literacy Using Interactive Simulations[J].Sustainability, 2022, 14.

[3] Vitorino E V .Indicators for Information Literacy in Brazil: virtues, trends and possibilities[J].perspectivas em ciencia da informacao, 2022, 27 (2):7-36.

[4] Ferguson J .Flipping the (COVID-19) Classroom: Redesigning a First-Year Information Literacy Program during a Pandemic[J].portal: Libraries and the Academy, 2023, 23:145 - 168.

[5] Carless D .From teacher transmission of information to student feedback literacy: Activating the learner role in feedback processes:[J].Active Learning in Higher Education, 2022, 23(2):143-153.

[6] Paul M , Deja M ,Magorzata Kisilowska-Szurmińska, Gowacka E, Marzena W, Wojciechowska M.Understanding information literacy among doctoral students: an ILDoc model and assessment tool[J].The Journal of Academic Librarianship, 2024, 50(2).

[7] Fei D , Erjun S .How the ICT Development Level Influences Students' Digital Reading Literacy: A Multi-level Model Comparison Based on PISA 2018 Data[J].Frontiers of Education in China, 2022, 17(2):151-180.

[8] Esfandiari R , Arefian M H .Developing collective eyes for Iranian EFL teachers' computer-assisted language assessment literacy through internet- based collaborative reflection[J].Education and Information Technologies, 2024, 29(8):9473-9494.

[9] Riithi C W , Kimani G .Assessment of Literacy Levels of Small-Scale Poultry Farmers for improved information behaviour and production: case of Kabaa location, Kenya[J].International Journal of Current Aspects, 2022.

[10] Ganesan P , Gunasekaran M .Assessment of information literacy skills and knowledge-based competencies in using electronic resources among medical students[J].Digit. Libr. Perspect. 2022, 38:444-459.

[11] Vianna B I , Caregnato S E .Institutional diagnosis models for implementation of Information Literacy programs in academic libraries[J]. perspectivas em ciencia da informacao, 2022, 27(2):242-267.

[12] Faber C J .Information Literacy Modules for First-Year Engineering Students[J].AEE Journal, 2022.

[13] Yu C , Xu W .Writing assessment literacy and its impact on the learning of writing: a netnography focusing on Duolingo English Testexaminees[J]. Language Testing in Asia, 2024, 14(1).

[14] Schmidt L A J , Deschryver M .The Role of Digital Application Literacy in Online Assessment.[J].Journal of Educational Technology Systems, 2022, 50.

[15] Stellwagen Q H , Rowley K L , Otto J .Flip This Class: Maximizing Student Learning in Information Literacy Skills in the Composition Classroom through Instructor and Librarian Collaboration[J].Journal of library administration, 2022.

[16] Current M D .Tracking student learning outcome engagement at the reference desk to facilitate assessment[J].Reference services review, 2023.

[17] Ke D D , Suzuki K , Kishi H , Kurokawa Y, Shen S. Definition and assessment of physical literacy in children and adolescents: a literature review[J]. Journal of Physical Fitness and Sports Medicine, 2022, 11(3):149-159.

[18] Xiao Y L, Sehn H R, Xu Y H, Yu T G, Zheng Y J, Xie H Z, Ting J. Water quality prediction in Minjiang River basin based on GBDT-LSTM[J]. Journal of Ecology and Environment,2024,33(04):597-606.

[19] Nie X ,Hong Y. Long-term power load forecasting based on WOA-VMD-GBDT[J]. Science and Technology Wind,2024,(12):70-72.

[20] Chen H, Shang B.. Mathematical modeling and performance analysis of paper mill beating degree GBDT[J]. Paper Science and Technology,2024,43(04):73-76.

[21] Emami H .Stock exchange trading optimization algorithm: a human-inspired method for global optimization[J].Journal of supercomputing, 2022(2) :78.

[22] He J, Wang W, Liu C, Su Y. Research on the prediction and control method of tunnel surface settlement based on GBDT-PSO hybrid algorithm[J]. Transportation World,2024,(13):181-183.

[23] Sun B L, Sun B W. Entropy optimization with improved SCA for image segmentation and its application to image recognition[J]. Computer Engineering and Design,2024,45(05):1516-1524.

[24] Kumar M, Kulkarni A J, Satapathy S C. Socio evolution & learning optimization algorithm: a socio-inspired optimization methodology[J]. Future Generation Computing System, 2018, 81:252-272.

[25] Askari Q, Saeed M, Younas I. Heap-based optimizer inspired by corporate rank hierarchy for global optimization[J]. Expert Systems with Applications, 2022, 161: 113702.

[26] Houssein E H , Saad M R , Hashim F A , Shaban H, Hassaballah H. Lévy flight distribution: a new metaheuristic algorithm for solving engineering optimization problems[J].Engineering Applications of Artificial Intelligence, 2020, 94: 103731.