IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 12 April 2024, accepted 3 May 2024, date of publication 23 May 2024, date of current version 3 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3404606

RESEARCH ARTICLE

Optimization Strategy of Classification Model Based on Weighted Implicit Optimal Extreme Learning Machine

NA ZHANG[1] AND XIAOFENG WANG[2]
[1]School of Computer and Artificial Intelligence, Henan Finance University, Zhengzhou 450046, China
[2]Graduate School of Business (GSB), SEGi University, Kota Damansara, Petaling Jaya 47810, Malaysia

Corresponding author: Xiaofeng Wang (wangxiaofeng2023zn@163.com)

This work was supported in part by the Science and Technology Research Project of Henan Provincial Department of Science and Technology: Research on Big Data Analysis and Mining Based on Deep Learning under Project 212102210501, in part by the Educational Teaching Reform and Practice Project of Henan Province: Construction and Practice of Financial Corpus Based on Industry-Education Integration in Finance and Economics Colleges under Project 2021SJGLX599, and in part by the Key Research Project of Higher Education Institutions in Henan Province: Key Technology Research on Multi-Source Information Fusion for Educational Big Data under Project 22A520017.

ABSTRACT Classification algorithms are one of the important research topics in the artificial intelligence, widely applied in various scientific and engineering fields. Extreme learning machine is a single hidden layer feed-forward neural network algorithm. Compared with traditional neural network models, the training speed and the generalization ability are also better. In terms of methodology, this study first innovatively improves the traditional Grey Wolf Optimization (GWO) algorithm to enhance its convergence and search ability. Specific improvement measures include implementing the reverse learning strategy to reduce the initial dependence of the algorithm on population distribution, and adding exploration perception strategy to enhance its global search ability by calculating heuristic factors, so as to identify the global optimal solution more effectively. The results showed that the improved W-DH-ELM model had excellent performance on multiple standard data sets. In particular, the average accuracy was more than 90%, which was significantly higher than other benchmark classification models. In terms of operation efficiency, the running time of the new model on different data sets was significantly reduced, accounting for less than 25%, and the lowest running time was only 4.89%. These experimental results verify the effectiveness of the introduced intelligent optimization algorithm in improving the performance of traditional ELM model without changing the original model structure. The improved W-DH-ELM model not only maintains the fast training performance of ELM, but also has higher accuracy and stability, which shows its superiority in dealing with complex classification tasks. In summary, the weighted two-hidden layer extreme learning machine optimized by the improved GWO proposed in this study has significant advantages in classification problems, providing a new perspective for future machine learning applications and research.

INDEX TERMS Two-hidden layer extreme learning machine, Grey wolf optimization algorithm, classification model, reverse learning, perception strategy.

I. INTRODUCTION

Artificial intelligence plays a crucial role in key areas such as agriculture [1], disease diagnosis [2], and logistics management [3]. Machine learning, a multi-disciplinary approach, has achieved significant results in various tasks [4], [5]. Classification is a hot research direction in machine learning. Many scholars have proposed many methods [6]. However, neural networks have some drawbacks, such as requiring continuous iteration and longer running time. To address these issues, a single Hidden Layer (HL) feed-forward neural network model called Extreme Learning Machine (ELM)

The associate editor coordinating the review of this manuscript and approving it for publication was Qichun Zhang.

is proposed, which has significant advantages in parameter setting [7], [8]. The operational efficiency refers to the resource utilization and execution efficiency exhibited by the model during task execution. ELM uses random initialization parameters to obtain output layer weights analytically, eliminating the need for parameter updates and greatly reducing the model running time, making it an efficient neural network model [9]. However, ELM still faces challenges in addressing classification issues, such as model uncertainty due to random parameter generation and the inability to guarantee optimal classification performance. To address these challenges, this research aims to improve the ELM classification algorithm using intelligent optimization algorithms, proposing an improved Grey Wolf algorithm optimized Weighted Two-hidden layer Extreme Learning Machine (WTELM) model, which is also the core technology of research.

Based on the WTELM, the HL is added on the basis of the traditional ELM. The weight factor is introduced to improve the ability of the model to capture the internal structure of data. To improve the performance of the model, the structure and parameters are adjusted by optimizing the strategy. Relying on TELM, the weighted concept is innovatively introduced to propose the WTELM model. Meanwhile, it is planned to improve the Grey Wolf Optimization algorithm (GWO) by changing the update method of convergence factors. The reverse learning and exploratory perception strategy are introduced to enhance the search ability of the optimal solution.

Through model improvement and algorithm optimization, the research improves the model performance in capturing the internal structure of data and its ability to search for optimal solutions, thus improving the overall performance. This research has made innovative achievements in improving the ELM model, designing the intelligent optimization algorithm, and combining the two, providing new ideas and methods for the development of related fields. The first part introduces the current research of ELM and intelligent optimization algorithms. The second part provides a detailed description for the WTELM proposed in the study and the improved model used for optimization. The third part introduces the experiments and the results. The fourth part summarizes the research and explains the shortcomings and future research prospects.

ELM has garnered interest from researchers. Its speed and efficiency have led to widespread use in fields such as fault and medical diagnosis, image and classification processing. A. Gumaei et al. proposed a hybrid feature extraction and regularization ELM for brain tumor classification. The method pre-processed brain images with min-max normalization to improve the contrast. The findings suggested that it outperformed current technologies. In a random retention experiment, the accuracy increased from 91.51% to 94.233% [10]. X. Zhao et al. proposed a fault diagnosis method for multi-channel motor rotor systems. This method achieved rapid fusion and intelligent diagnosis of multi-channel data through the multi-manifold deep ELM algorithm. The proposed MDELM had higher efficiency [11]. F. T. Al Thief et al. proposed a sound pathology detection and classification system. The system used an online sequential ELM model as a classifier and Mel frequency coefficients as feature extraction. According to the findings, the model could achieve the highest accuracy of 91.17%. The precision reached 94%, and the recall rate reached 91%. In addition, the F-value, G-mean, and specificity of the model were 87%, 87.55%, and 97.67%, respectively [12]. The shortcomings of traditional ELM in wind power prediction had low prediction accuracy and model instability. Therefore, N Li et al. proposed a kernel ELM based on differential evolution and cross validation optimization to predict short-term wind power generation. According to the findings, the convergence speed was twice that of the genetic algorithm. Differential evolution method improved the model accuracy by 8.34% [13]. F. Zhang and other scholars developed an ELM model using voxel morphology measurement technology based on magnetic resonance imaging to distinguish Alzheimer's patients and normal control groups. The classification accuracy of the ELM model was as high as 0.96. The results of the other three methods were 0.82, 0.79, and 0.75, respectively. It performed best in distinguishing Alzheimer's patients, which was helpful for disease diagnosis [14]. Shariati et al. proposed a mixed model combining ELM and GWO for the concrete compressive strength prediction, but this method would lead to poor prediction results due to over-fitting in some cement materials [15]. Aiming at the low efficiency of ELM in processing large-scale data and the uncertainty in HL weights and bias configurations, M. A. Salam proposed a hybrid ELM algorithm combined with swarm intelligent optimization technology. The research proved that its accuracy was excellent [16]. Aiming at the serious consequences of open-pit blasting, C. Li et al. used the swarm intelligence optimization algorithm ELM model for prediction. The performance of the ELM model was improved. The experimental results proved its excellent prediction accuracy [17]. K. Roushangar proposed a nuclear ELM model based on GWO to improve the prediction accuracy of the water discharge coefficient to solve the poor accuracy. The research proved that the model had excellent performance compared with Support Vector Machine (SVM) and Gaussian Process Regression (GPR). However, there were still efficiency problems [18]. Aiming at the ECG signal separation problem in the diagnosis of heart disease, A. Diker et al. proposed the differential evolution algorithm combined with hidden neurons in ELM to solve the accuracy problem of ELM classification. The experimental results showed that the accuracy reached 83.12%, which still had a large optimization space [19].

Intelligent optimization algorithms imitate the behavior of biological groups such as predation and hunting to find the optimal solution. It can be searched by imitating the population characteristics of different organisms [20]. The research, improvement, and application of intelligent optimization algorithms have also been popular. W. Liu et al. proposed a novel Particle Swarm Optimization (PSO) algorithm. The newly proposed adaptive weighted strategy took into account both the distance from the particle to the global optimal position and the distance from the particle to the individual optimal position. According to the findings, it significantly improved the convergence speed of the particle swarm optimizer and outperformed the currently popular PSO [21]. X.F. Song et al. proposed a three-stage hybrid feature selection algorithm based on correlation guided clustering and PSO. The aim was to address the dimensional and high computational cost constraints of high-dimensional feature selection problems. It could obtain a good subset of features with the lowest computational cost [22].

Previous studies indicate that ELM has a rapid training process and robust generalization capability, which can be applied in various fields. The GWO algorithm is notable for its high performance and straightforward parameter configuration, making it a preferred choice for engineering and optimization tasks. Practical applications frequently necessitate customized enhancements. Consequently, the WTELM model is introduced, integrating weights into TELM. The model employs reverse learning and exploratory strategy to improve the GWO. The refined algorithm optimized WTELM, generating a classification model with enhanced stability, precision, and speed.

II. AN IMPROVED GREY WOLF OPTIMIZATION ALGORITHM FOR OPTIMIZING WEIGHTED DOUBLE HIDDEN LAYER LIMIT LEARNING MACHINES

Based on GWO's excellent adaptability to classification problems, it is chosen as the basic algorithm. The motivation of combining ELM and GWO. GWO uses global search capability and excellent adaptability to improve the initial parameter selection of ELM in classification problems, thus enhancing the learning efficiency and classification accuracy of the model. The reverse learning initialization is used to improve its convergence accuracy and convergence speed. The heuristic perception strategy model is introduced to solve the problem of GWO falling into local optimal solution during the iteration process. Based on the improved GWO, the weighted concept is introduced to improve the classification performance of the model.

A. IMPROVED GREY WOLF OPTIMIZATION ALGORITHM

The GWO is a heuristic optimization algorithm that simulates the hierarchy and hunting behavior of grey wolves [23]. Heuristic perception strategy is a technique used in optimization algorithms, which guides the search process based on heuristic rules to explore the solution space more efficiently and find the global optimal or approximately optimal solution. When it comes to parameter optimization of machine learning models, heuristic perception strategies often draw inspiration from phenomena or biological behaviors in nature, such as natural selection and genetic mechanisms in genetic algorithms, or bird hunting behavior in PSO. In the classification problem, the social hierarchy of GWO maintains a balance between exploration and utilization. Compared with traditional classification algorithms such as SVM and Naive Bayes (NBM), Wolf pack cooperation strategy can help the algorithm jump out of the local optimal and improve the probability of finding the global optimal solution. The simple rules and dynamic adjustment strategy make the algorithm adaptable and easy to implement. According to the hunting process of grey wolves, the hunting behavior can be summarized into three steps. The first is to track and approach the prey. During the tracking and approaching prey, the grey wolf updates the position by calculating the distance $\vec{D}$ in the grey wolf and the prey, as shown in formula (1) [24].

$$\begin{cases} \vec{D} = \left\| \vec{C} \cdot \vec{X}_p(t) - \vec{X}(t) \right\| \\ \vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \end{cases} \tag{1}$$

In formula (1), t is the current iteration. $\vec{X}$ and $\vec{X}_p$ are position vectors of prey and grey wolf. $\vec{A}$ and $\vec{C}$ are coefficient vectors, respectively, as shown in formula (2).

$$\begin{cases} \vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \\ \vec{C} = 2\vec{r}_2 \end{cases} \tag{2}$$

In formula (2), $\vec{a}$ is the convergence factor. Its value decreases linearly within the [0, 2] range during the iteration process. $\vec{r}_1$ and $\vec{r}_2$ are random numbers within [0, 1]. The second step of grey wolf hunting is to chase and surround the prey. The grey wolf population consists of four levels, from highest to lowest being α wolf, β wolf, δ wolf, and ω wolf. In the actual hunting process, the α wolf cannot determine the specific location. Therefore, the location of the α wolf is the potential optimal solution location, followed by the β wolf and δ wolf. When chasing prey, the top three wolves guide the action of ω wolf and update the position of individual grey wolves. The process of grey wolves surrounding their prey is shown in Figure 1.

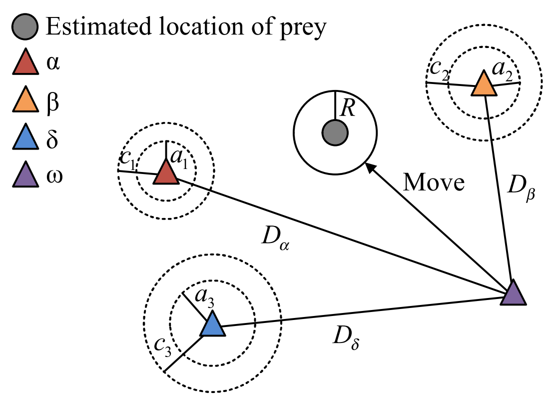

FIGURE 1. Schematic diagram of the process of grey wolves surrounding prey.

After tracking and surrounding the prey, formula (3) is used to adjust the position of the grey wolf and prey, and then launch an attack on the prey.

$$a = 2 - \frac{2t}{T} \tag{3}$$

In formula (3), T is the maximum iteration. An increase in the a value indicates that the grey wolf is moving away from the prey for a global search. If the a value decreases, it indicates that the grey wolf is approaching the prey for local search. The strict social hierarchy system in the grey wolf population is used as a biological prototype. The hunting process, group coordination, and other population behaviors of the grey wolf are described through mathematical language. In the improved GWO algorithm, α wolf still holds the highest social status. It determines the actions of the entire wolf population. The other grey wolves are all under control. However, as a qualified leader, wolves also actively listen to the opinions of other grey wolves. β wolf ranks second and needs to follow the command of α wolf. The opinions of α wolf are conveyed to other grey wolves. It can also direct the activities of δ and ω. When α wolf ages or experiences accidents, β wolf becomes the best candidate as the second leader. δ wolf has a lower status and operates under the control of α wolf and β wolf, but it can guide the behavior of ω wolf. ω is the lowest level member and must obey the commands of α, β, and δ, without any command authority. Due to the position at the bottom of society, it has become a target for α wolf, β wolf, and δ wolf to vent their emotions, making important contributions to maintaining the harmony and stability of the population. When grey wolves hunt, they first need to track their prey and gradually approach their location. The improved grey wolf position update is shown in formula (4).

$$\begin{cases} \vec{D}_{new} = \left\| \vec{C} \cdot \vec{X}_{pnew}(t) - \vec{X}_{new}(t) \right\| \\ \vec{X}_{new}(t+1) = \vec{X}_{pnew}(t) - \vec{A} \cdot \vec{D}_{new} \end{cases} \tag{4}$$

The coefficient vectors $\vec{A}$ and $\vec{D}_{new}$ remains unchanged. The prey encirclement is commanded by three headed wolves, namely α wolf, β wolf, and δ wolf. However, they actually do not know the specific location of their prey. Therefore, in the improved GWO, the location of the three headed wolves is considered as a potential location for prey. Assuming that three wolves have a better understanding for the location information, they can guide other prey to update and iterate. The position $\vec{X}_{new}(t+1)$ of the current solution is shown in formula (5).

$$\begin{aligned} &\vec{X}_{new}(t+1) \\ &\quad = (f_\alpha/f) \cdot \vec{X}_{new\_1} + (f_\beta/f) \cdot \vec{X}_{new\_2} + (f_\delta/f) \cdot \vec{X}_{new\_3} \end{aligned} \tag{5}$$

In formula (5), $\vec{X}_{new\_1}$, $\vec{X}_{new\_2}$, and $\vec{X}_{new\_3}$ are updated positions of α wolf, β wolf, and δ wolf. f_α, f_β and f_δ are the fitness values calculated by the objective functions based on the positions of α wolf, β wolf, and δ wolf. The sum of the three is f. In the improved GWO algorithm, when prey surrounded by grey wolves no longer moves, grey wolves begin to attack their prey. When prey is caught, the grey wolf terminates the hunting activities. When attacking prey, the size of $\vec{A}$ is used to determine whether the grey wolf launches an attack on the prey. The decreasing method and value range of the convergence factor $\vec{a}_{new}$ determine the size of $\vec{A}$. The value range of $\vec{A}$ is between $(-2a, 2a)$. The $\vec{a}_{new}$ is shown in formula (6).

$$\vec{a}_{new} = 1 + \cos(\pi \cdot t/T) \tag{6}$$

The convergence time of the GWO algorithm in solving task scheduling is related to the distance between the grey wolf leadership and the optimal solution. If the adaptability of the leadership is high enough, that is, if it is close to the optimal solution individual, it indicates that the convergence speed of the algorithm is faster [25], [26]. The classical GWO algorithm generates initial positions for each dimension through random functions in the initial stage. This results in the convergence speed being affected by the initial position of the grey wolf. The quality of the initial position determines the convergence accuracy and speed. Therefore, the reverse learning is used to initialize the population in the initial stage. It can enable grey wolves to undergo manual intervention in the initial stage, making a purely random event controllable to improve the quality of the initial population. At the same time, the GWO algorithm falls into a local optimal solution during the iteration process. Therefore, a further OTGWO algorithm is proposed. The exploratory perception strategy model proposed in the study is shown in Figure 2.

In Figure 2, the OTGWO algorithm generates h exploratory perception factors q around all grey wolves during the iteration process. The size of q depends on the problem, initial population, and iteration. When the problem size is the same, a smaller initial population indicates a larger exploratory perception factor value q. As the iterations increase, the leadership ability of grey wolf tends to stabilize, making it easier to find the global optimal solution in the connection space. At this point, the number of q should decrease. In summary, the initial value of q is inversely proportional to the initial populations. It is also linearly decreases with the increase of iteration times. The process of OTGWO algorithm is shown in Figure 3.

By introducing heuristic sensing factors, the global search capability of GWO can be significantly enhanced. Compared with the standard GWO algorithm, the heuristic perception factor provides an additional exploration mechanism, allowing the algorithm to find the optimal solution in a wider search space. Specifically, when the fitness of the solution generated by a single grey wolf exceeds its current position according to the temptation perception factor. The temptation position will replace the original position, thus guiding the grey wolf to move to a potentially better area. Compared with the gray wolf iteration in the standard GWO algorithm, probing the perception factor can provide a wider search space. When the fitness of the j-th exploratory perception factor produced by the i-th grey wolf is better than that of the grey wolf itself, the exploratory perception factor replaces the original grey wolf. When all exploratory perception factors are processed,

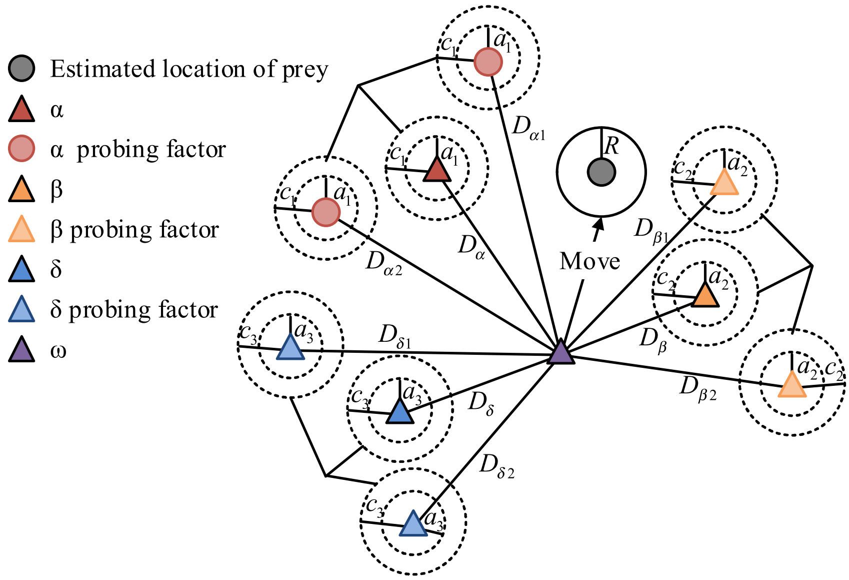

FIGURE 2. Schematic diagram of exploring the perception strategy model.

the fitness of all grey wolves is recalculated and the α, β, and δ wolves with leadership abilities are re-selected.

B. OPTIMIZATION WEIGHTED DUAL HIDDEN LAYER ELM CLASSIFICATION MODEL BASED ON OTGWO ALGORITHM

To improve the classification performance and generalization ability of ELM algorithm, the weighted idea is further introduced into ELM algorithm. This improvement is mainly for the output of HL neurons in the ELM algorithm. Each output is introduced with a weighted factor to enhance or weaken the contribution of certain neurons, thereby better capturing and processing the complex relationships of input data. The difference between ELM and traditional neural networks is that the parameters from the input to the hidden in the ELM are randomly initialized. The weights from the HL to the OL are analytically calculated. There are N different samples $\left\{(x_i, y_i) \left| x_i \in R^d, t_i \in R^m \right.\right\}_{i=1}^{N}$. $x_i = [x_{i1}, x_{i2}, \cdots, x_{id}]^T$ is the i-th training sample. The dimension is d. $t_i = [t_{i1}, t_{i2}, \cdots, t_{im}]^T$ is the expected output of i. Therefore, the output matrix H of the ELM model with L HL nodes is shown in formula (7).

$$H = \begin{bmatrix} h(x_1) \\ h(x_2) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} g\left(w_1^T x_1 + b_1\right) & \cdots & g\left(w_L^T x_1 + b_L\right) \\ g\left(w_1^T x_2 + b_1\right) & \cdots & g\left(w_L^T x_2 + b_L\right) \\ \vdots & \ddots & \vdots \\ g\left(w_1^T x_N + b_1\right) & \cdots & g\left(w_L^T x_N + b_L\right) \end{bmatrix} \tag{7}$$

In formula (7), (w_i, b_i) is the parameter of the i-th neuron in the hidden. $W = (w_1, w_2, \cdots, w_L)$ and $b = (b_1, b_2, \cdots, b_L)$ are parameters that act on the input layer of the HL. $g(\cdot)$ is the activation function of HL neurons. The commonly used activation functions include Radbas function, Sine function, Sigmaid function, and Hard limit function. The training goal of ELM is to train model parameters to approximate the training samples. The objective function $L_{Minimize}^{ELM}$ is shown in formula (8).

$$L_{Minimize}^{ELM} = \frac{1}{2} \left\| H\beta - T \right\|_2^2 \tag{8}$$

In formula (8), β is the output weight. T is the objective matrix of the training set. When the activation function is infinitely differentiable, the HL output matrix H is fixed for the given parameters w_i and b_i. According to the relationship between the HL neurons L and the training samples N, the output weight β is solved in two cases. When $L = N$, the output matrix H of the HL is an inverse matrix. The output weight β can be directly obtained by solving the inverse matrix. When the two are not equal, the least squares method is used to solve the output weight β. To alleviate the shortcomings of ELM in feature extraction while retaining the fast learning ability, TELM is proposed. The TELM is displayed in Figure 4. Two-hidden ELM improves the learning ability and generalization performance of the original ELM model by introducing additional HLs. This enhanced network structure allows the model to capture more complex and advanced features in the input data, especially when dealing with high-dimensional data or complex data structures. Each HL can be considered as a nonlinear converter. The first HL captures the primary features of the original data, while the second HL further abstracts and combines these features to form a more abstract high-level feature representation.

In the TELM model, the weight and bias parameters of the input layer to the first HL are randomly initialized. Then the output of the first HL is calculated based on these parameters and input data. Based on the output of the first HL, the model can calculate the weight of the second HL, and further get the weight from the second HL to the output

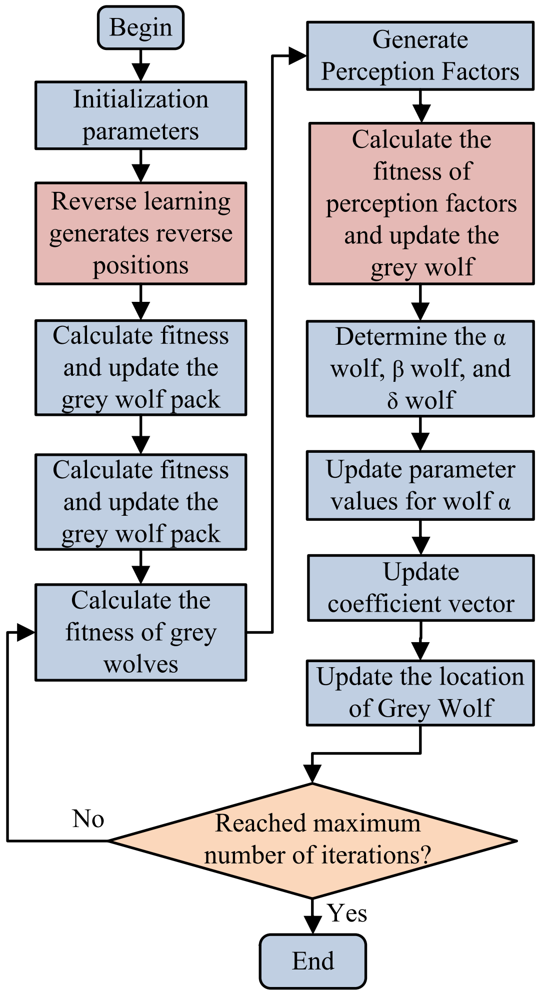

FIGURE 3. Flowchart of OTGWO algorithm.

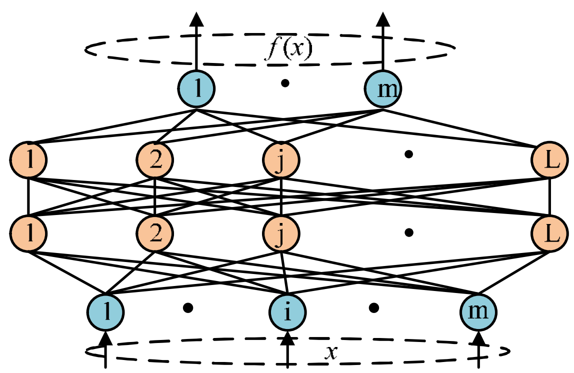

FIGURE 4. Structure diagram of TELM.

layer. This hierarchical structure helps the model capture more complex features or patterns in the data, theoretically improving the performance of ELM in various tasks. When it comes to a large number of parameter optimizations and complex network structures, powerful computing power is often required, which may be provided by high-performance CPUs or GPUs. At the same time, model training and data processing also require sufficient memory and storage space to ensure that the data can be efficiently loaded and the model state can be saved. In addition, the optimized machine learning libraries can increase efficiency and reduce resource consumption. In practical applications, resource allocation needs to be balanced according to the application scenario and performance requirements of the model to achieve the best balance between efficiency and cost. The main idea of TELM is to add a HL structure on top of ELM. The constraint condition is that the new HL nodes are the same as the original HL [27]. When improving the global search capability, this method may introduce certain biases, such as biases in certain regions of the solution space during the search process, reducing population diversity without proper control, or increasing the risk of converging to local optimal solutions. The parameter selection of reverse learning is also an issue that needs careful consideration, as inappropriate parameter settings can introduce additional biases. Therefore, in practical applications, the reverse learning strategy needs to be carefully adjusted to ensure that the advantages it brings outweigh any potential biases. There are two main purposes. One is to avoid various problems caused by the excessive number of HL nodes, such as over-fitting and increased computational complexity. The other is to effectively reduce the redundant and invalid nodes in the HL parameters randomly set by ELM [28]. In the TELM, the parameters from the input layer to the first HL are randomly generated. Then, the weights from the first HL to the second HL and from the second HL to the OL are calculated. Any N independent training samples (x_i, t_i) $(i = 1, 2, \cdots, N)$ and $2L$ HL nodes are given. The weight W and bias B from the input layer to the first HL are randomly initialized. The output matrix of the first HL is shown in formula (9).

$$g\left(X_{IE}W_{IE}\right) = H \tag{9}$$

In formula (9), $W_{IE} = [BW]$, and $X_{IE} = [1X]$. From this, the expected output $H_1 = T\beta^+$ of the second HL can be obtained. β^+ represents the Moore Penrose generalized inverse matrix of the output weight β. The weight of the second HL is shown in formula (10).

$$W_{HE} = H_E^+ g^{-1}\left(H_1\right) \tag{10}$$

In formula (10), $H_E = [1H]$. Therefore, the actual output of the second layer can be shown in formula (11).

$$H_2 = g\left(H_E W_{HE}\right) \tag{11}$$

According to the Moore Penrose generalized inverse matrix H_2^+ of matrix H_2 and the connection weight β_{new} between the second HL and the OL, the output $f(x)$ of the TELM model can be obtained, as illustrated in formula (12).

$$f\left(x\right) = H_2\beta_{new} \tag{12}$$

Based on the ELM, the TELM model adds an additional HL. Traditional ELM models require more HL nodes to ensure its effectiveness. The proposed TELM model overcomes this limitation. Excellent model performance can still be achieved with fewer HL nodes. However, the TELM model does not consider data structure and distribution characteristics. Therefore, based on TELM, the weighted idea is introduced to propose WTELM. According to the data distribution, each group of data is given different weights. The weight calculation is shown in formula (13).

$$w_j = 1 - \frac{M\left(t_j\right)}{N}, j = (1, N) \tag{13}$$

In formula (13), N is the sample in the data set. $M\left(t_j\right)$ represents the sample in class j. After introducing the weighted idea, the data is shown in formula (14).

$$Data_w = w_j \cdot N_j, j \in (1, N) \tag{14}$$

In formula (14), N_j represents the j-th data. w_j represents the weight of the j-th data. The improved GWO is creatively applied to optimize the parameters of WTELM. Theoretically, the optimized weight can more effectively capture the features and internal structure of the input data, thereby improving the generalization ability of the model to new data. This optimization typically utilizes regularization techniques to reduce the over-fitting, adjust the learning rate to balance the convergence speed and stability, and gradually improve the prediction accuracy of the model during the training process. A classification model based on IGWO-WTELM is proposed. The calculation process of the model is shown in Figure 5.

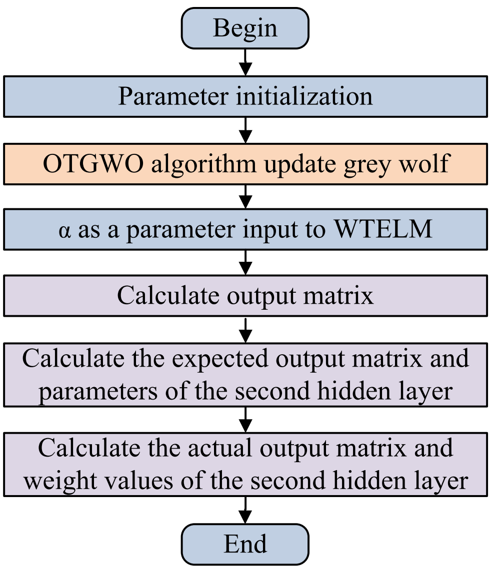

FIGURE 5. **Flowchart of IGWO-WTELM classification model.**

Firstly, parameters are initialized. Then the OTGWO updates the individual's position in the grey wolf pack to find the optimal parameters of the WTELM model. The optimal solution is entered into WTELM as a parameter. The model calculates the output matrix accordingly, and then estimates the desired output matrix and parameters of the second HL and adjusts them to fit the target. Finally, the actual output matrix and weight values are calculated according to the updated second HL parameters. In the model decision-making, the difficulties mainly include randomness of initial parameter setting, criticality of optimization algorithm selection, high requirement of optimization efficiency, risk of over-fitting, complexity of parameter adjustment and uncertainty of model generalization ability. These challenges together affect the effectiveness of model training and the quality of final decisions. The WTELM has two HLs. Therefore, the model adjusts the parameters of the second HL to better fit the target, which may involve adjusting the weights to reduce the loss function value. Once the second HL parameters are optimized, the model uses the updated parameters to calculate the actual output matrix. In a classification task, this is equivalent to giving a predicted score for each category based on updated weights. At the end of the process, the model evaluates classification performance by comparing the predicted output with the real label. The feedback is used to further iterate and optimize the location of the grey wolf population. To evaluate the classification efficiency of the WTELM optimized by intelligent optimization algorithms, the ratio $Time\_Percent_{ji}$ of the time spent on each model to the running time of all models is used as the evaluation indicator, as shown in formula (15).

$$\begin{aligned} &Time\_Percent_{ji} \\ &\quad = Module\_time_{ji}/All\_time_j \times 100\% \end{aligned} \tag{15}$$

In formula (15), $Module\_time_{ji}$ is the running time of the i-th model on the j-th data set. All_time_j is the running time sum of all models on the j th data set. Figure 6 shows the pseudo-code of the proposed the method framework.

III. CLASSIFICATION MODEL PERFORMANCE BASED ON IGWO-WTELM

To verify the effectiveness of the proposed WTELM model and the impact of intelligent optimization algorithms on model classification performance, different experiments are designed. The comparative models used in the experiment include ELM, RELM, and TEL. PSO, MFO, GWO, and OGWO algorithms are used to optimize the WTELM model. The classification accuracy, running time, stability, and convergence are compared to select the model with the best classification performance.

A. EFFECT OF CLASSIFICATION MODEL BASED ON WTELM

The effectiveness validation experiment of the WTELM model uses five sets of classification data. The selected data set is all from the University of California, Irvine (UCI) database. The UCI machine learning database is a public data set resource widely used for research and teaching in machine learning and data mining. The Quantitative Structure-Activity Relationships (QSAR) data set is an important resource in the environmental science. It is used to predict the biodegradability of compounds. The Glass Identification data set contains information about the chemical composition of glass samples, which is often used to distinguish glass types. The Bupa data

```
Initialize the grey wolf pack (a set of solutions, where each wolf represents a set of WTELM parameters)
Initialize WTELM parameters (weights and biases)
Evaluate the fitness of the wolf pack (using the classification performance of WTELM as the fitness)

while (not reached max number of iterations) do
   for each wolf in the pack do
      Evaluate its fitness using WTELM
      Update the wolf's personal best solution if necessary
   end for

   Identify and record the best three solutions (alpha, beta, delta)

   for each wolf in the pack do
      Calculate new position based on the positions of alpha, beta, and delta
      Ensure the new position is within the search space
      Evaluate the fitness of the new position using WTELM
   end for

   Update the positions of alpha, beta, and delta if better solutions are found
   Implement any additional optimization strategies (like escape behavior) if necessary

   if (improvement in fitness is marginal) then
      Adjust parameters or implement diversity enhancement strategies
   end if
end while

Output the optimal solution (the position of the alpha wolf corresponding to WTELM parameters)
```

FIGURE 6. Schematic of the pseudo-code of the proposed method of the study.

TABLE 1. Experimental data set.

Number	Data set name	Number of categories	Characteristic dimension	Data volume	Label distribution
1	QSAR Biodegradation	2	41	1055	(699, 355)
2	Glass Identification	6	9	214	(6976, 17, 13, 9, 29)
3	Bupa	2	6	245	(144, 200)
4	CMC	2	9	1473	(844, 628)
5	Prnn_fglass	6	9	214	(13, 9, 17, 29, 69, 76)

set focuses on liver disease studies, particularly those related to alcohol consumption, which provides some biochemical test results for patients. The Contraceptive Method Choice (CMC) data set is based on survey data from women in India, covering the relationship between contraceptive choice and socioeconomic status. The Prnn_fglass data set may be a variant of glass recognition data used for pattern recognition and neural network algorithm research. The training set and testing set are randomly selected in a 8:2 ratio. The range of HL nodes for ELM, RELM, TELM, and WTELM models is $(10, 500)$. The parameter range from input layer to HL is $(0, 1)$. In the RELM model, the range of regularization coefficients is $\left(2^{-50}, 2^{50}\right)$. Each group of data is performed to 50 experiments. The accuracy and the time required to run once are recorded. The detailed information of the data set used in the study is shown in Table 1.

The classification accuracy and running time results obtained by each model on five datasets are shown in Figure 7. From Figure 7 (a), WTELM performed the best in accuracy on all datasets, with values of 0.97, 0.81, 0.85, 0.98, and 0.76, respectively. In contrast, the other three models performed similarly on the QSAR Biodegradation data set, all ranging from 0.84 to 0.88. However, there was a significant difference in performance on the other four datasets. The performance of ELM was relatively poor. From Figure 7 (b), WTELM had no significant advantages in running time compared with ELM. However, compared with RELM and TELM, WTELM had significant advantages in running time. Taking the QSAR Biodegradation data set as an example, RELM took 2852.11s, while TELM was 90.26s. The WTELM was 89.09s, which was much higher than comparative models. Overall, although the running time of WTELM is not the shortest, it performs excellently in classification accuracy. This means that although it takes slightly more time, WTELM can obtain more accurate classification results. Therefore, WTELM performs the best on these five datasets.

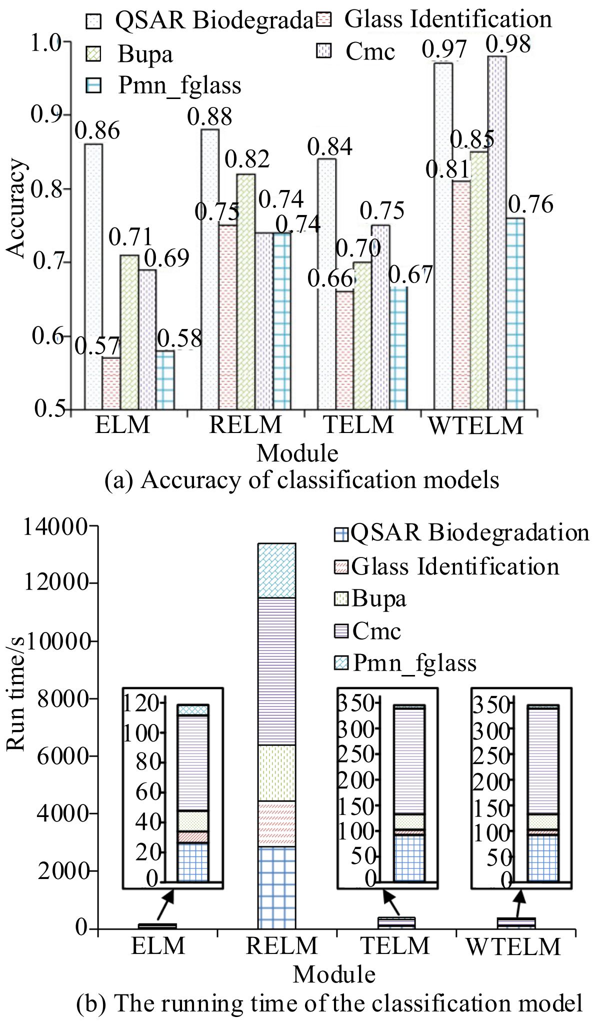

FIGURE 7. Accuracy and running time results of classification models.

ELM, RELM, TELM, and WTELM use randomly generated input layer parameters that have a certain impact on model stability. Even if the same initial HL nodes are set in each experiment, the classification results of the model still differ. To evaluate the stability of the model, a box plot is drawn based on the 50 accuracy experiments. The classification stability of model is evaluated according to accuracy fluctuations. In Figure 8, the vertical axis represents the accuracy, and each box represents a different method. From Figure 8, the accuracy of the WTELM model was always controlled within a small range. Although the accuracy fluctuation was not guaranteed to be minimal on all datasets, the WTLM model had better stability.

Based on the impact of the HL nodes on the classification accuracy, the complexity of the model is explored. One of the parameters to be explored in the RELM model is the regularization coefficient. Therefore, no comparison is added. From Figure9 (a), a quantitative analysis of the QSAR Biodegradation data set revealed that the following specific trends in model accuracy with varying numbers of HL and TELM models maintain a relatively stable accuracy, with a coefficient of variation (CV) of less than 2% as the number of nodes range from 20 to 100. The experiment depicted in Figure 9 (b) for the Bupa data set showed a different trend. As the number of HL nodes increased from 5 to 50, the accuracy of the ELM, TELM, and WTELM models decreased by 8%, 6%, and 4.5%, respectively. For the Prnn_fglass data set, as shown in Figure 9 (c), the relationship between the number of HL nodes and classification capability demonstrated a plateau effect. When the number of HL nodes ranged from 10 to 50, the ELM, TELM, and WTELM models showed a marginal variance in accuracy, with a CV of around 1.5%. Overall, the WTELM model outperforms other models in maintaining high classification performance in various HL node configurations. Quantitatively, the WTELM model consistently achieved a classification accuracy 3-5% higher than the ELM model and 1-3% higher than TELM model under identical HL node conditions.

The Ablation experiment is used to test the effectiveness of the improvements made in the study. The test results are shown in Table 2. From Table 2, the combined optimization strategy added by IGWO-WTELM was more effective than all single optimization strategies. WELM and TELM models performed better than ELM, while WTELM models performed better than WELM and TELM. IGWO-WTELM models performed best. In contrast, WELM model optimized the weight distribution in the training process by introducing weight adjustment, thus improving the generalization ability of the model. The TELM model improves its adaptability and prediction accuracy to time series data by considering time factors. When the two models are combined into WTELM, the performance is further improved, which indicates that the dual optimization strategy based on the weight and time can complement each other and improve the overall performance of the model.

B. CLASSIFICATION MODEL PERFORMANCE BASED ON IGWO-WTELM

To verify the optimization ability of the proposed OTGWO and analyze the impact of using intelligent optimization algorithms on model classification performance, PSO, MFO, GWO, OGWO, and OTGWO are used to optimize the parameters of WTELM, respectively. PSO-WTELM, MFO-WTLM, GWO-WTELM, and IGWO-WTELM classification models are established. Then, the classification performance

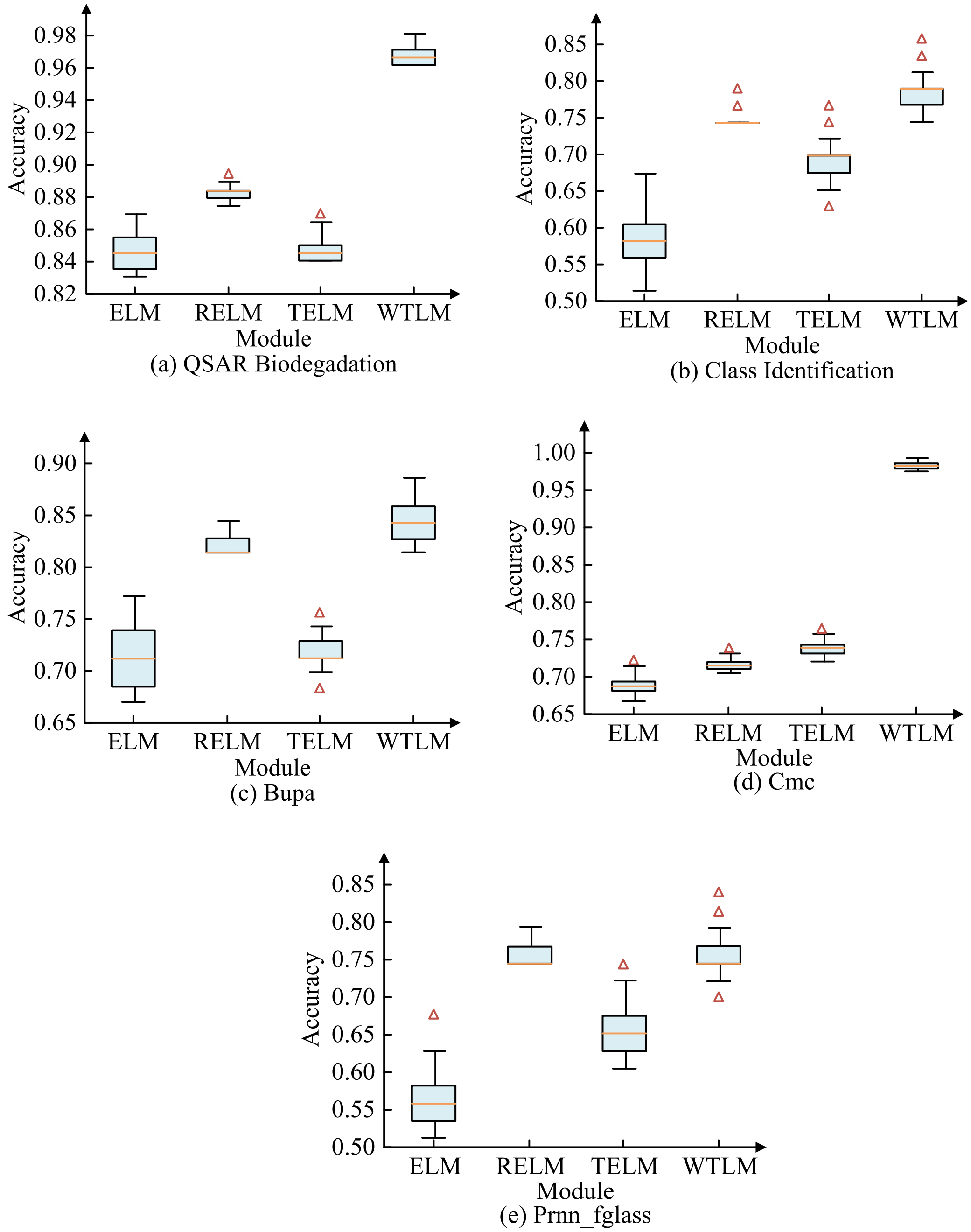

FIGURE 8. Comparative analysis of classification model stability.

TABLE 2. Results of ablation experiments.

Experiment No.	Accuracy rate (%)				
	IGWO-WTELM	WTELM	WELM	TELM	ELM
1	97.8	81.2	76.4	75.4	62.1
2	98.1	82.1	73.2	76.1	55.6
3	97.6	80.1	75.1	74.2	56.7
4	98.2	82.3	74.2	75.2	58.4
5	97.2	81.4	72.5	74.5	60.1
Average	97.8	81.4	74.3	75.1	58.6

of each model is analyzed from classification accuracy, stability, convergence, and running time. During the experimental analysis, four sets of data are selected for modeling, all of which are down-loaded from the UCI database. The data

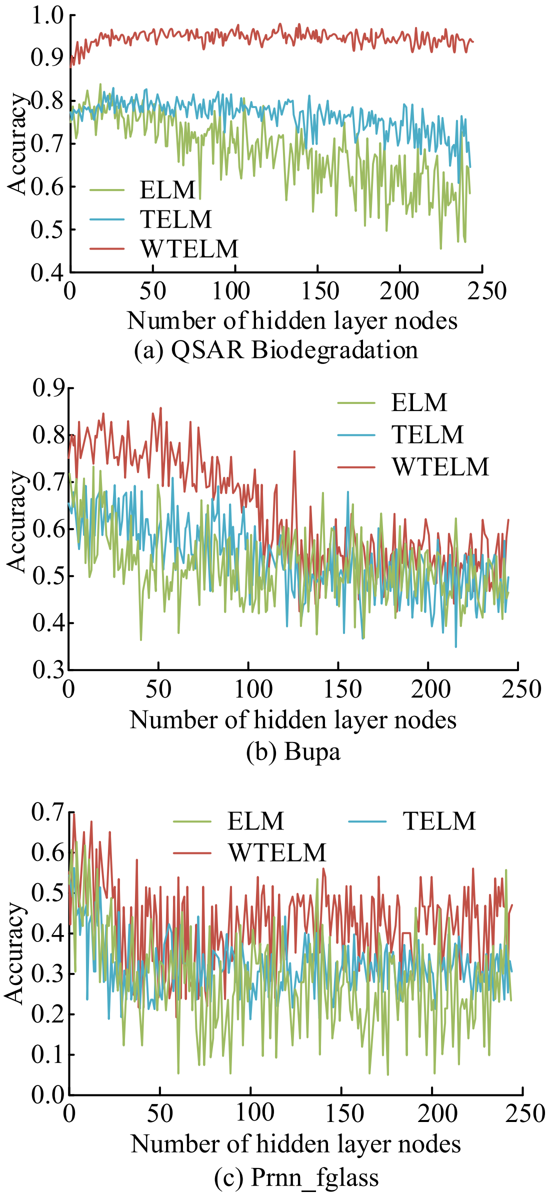

FIGURE 9. Comparative analysis of classification model complexity.

set includes QSAR Biodegradation, Bupa, Prnn_fglasss and Dermatology, as shown in Table 3.

During the experiment, all datasets are randomly selected in a 8:2 ratio for training and testing. The training set trains the model parameters, while the testing set tests the model classification performance. The range of HL nodes in the WTELM model is (10, 500). The population is (100, 300). The maximum iteration is (100, 500). The inertia factor of the PSO algorithm is 0.8. The two acceleration constants are 2 and 0. The two perturbation factors are random numbers within (0, 1). 10 experimental analyses are conducted on four data sets using different classification models. The accuracy results obtained are displayed in Figure 10. From Figure 10 (a), the accuracy of the classification model using intelligent optimization algorithms was higher than that of the WTELM model in most cases. From Figure 10 (b), on the Bupa data set, the OTGWO-WTELM model had the highest

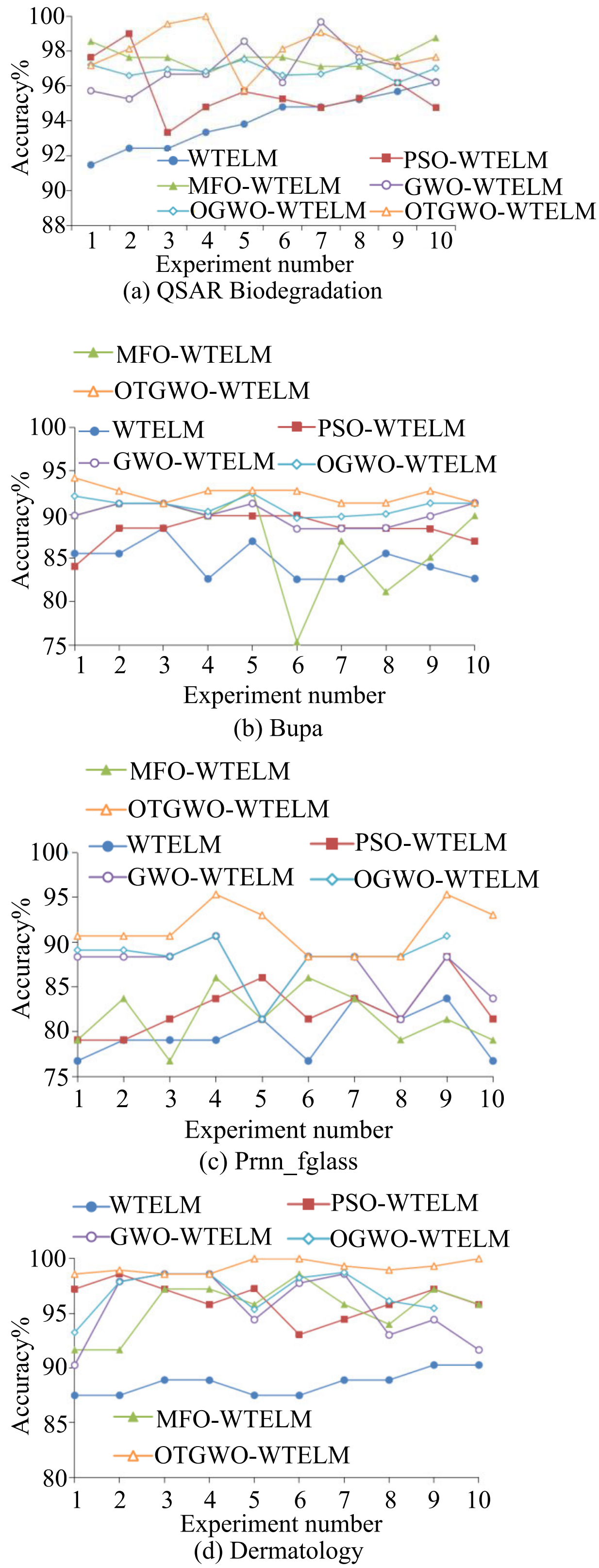

FIGURE 10. Comparison of accuracy of various models on different datasets.

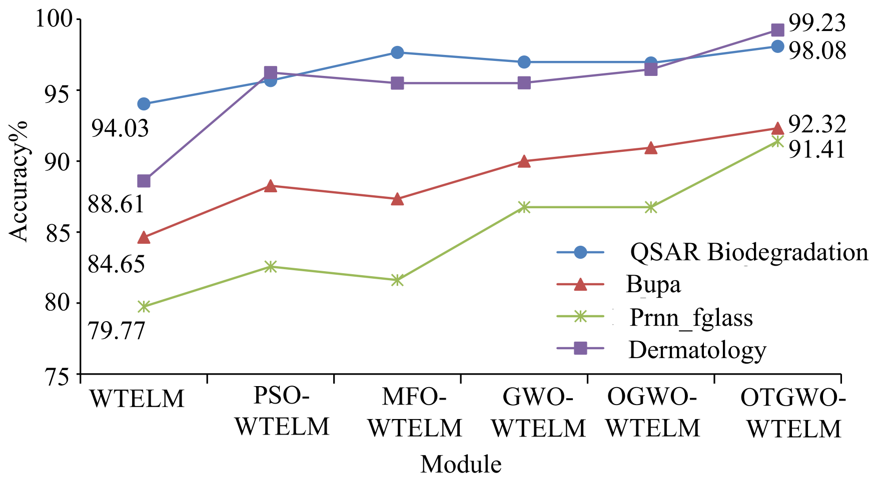

FIGURE 11. Comparison of average accuracy of various models on different datasets.

TABLE 3. Experimental data set.

Number	Data set name	Number of categories	Characteristic dimension	Data volume	Label distribution
1	QSAR Biodegradation	2	41	1055	(699, 355)
2	Bupa	2	6	245	(144, 200)
3	Prnn_fglass	6	9	214	(13, 9, 17, 29, 69, 76)
4	Dermatology	4	19	358	(247, 17, 59, 34)

TABLE 4. Independent data set test results.

Models	Fashion-MNIST		CIFAR-10		Iris Flower data set	
	Accuracy (%)	RMSE	Accuracy (%)	RMSE	Accuracy (%)	RMSE
IGWO-WTELM	97.8	0.012	98.1	0.024	98.2	0.027
FCM–WOA–ELM–GMM	90.7	0.156	91.2	0.172	90.3	0.184
CS-ELM	86.4	0.289	85.3	0.318	86.8	0.297
PCA-SDWPSO-ELM	80.1	0.417	78.2	0.432	79.6	0.421

accuracy, approaching over 92%. The accuracy of other models was generally between 80% and 90%. In Figure 10 (c), for Prnn_fglass data set, almost all models had an accuracy of over 80%. OTGWO-WTELM had the highest accuracy, reaching over 90%. From Figure 10 (d), almost all models had an accuracy of over 85% on the Dermatology data set. The accuracy of OGWO-WTELM and OTGWO-WTELM models exceeded 98% in almost all experiments. This indicates that the optimization algorithm introduced by reverse learning and exploratory perception has a high improvement effect on the classification accuracy of the WTELM. In summary, there are differences in the performance of different classification models on different datasets. Especially on some datasets, the WTELM model using GWO algorithm and improved GWO algorithm achieve better accuracy. These results indicate that intelligent optimization algorithms are crucial for improving the performance of classification models.

To evaluate the performance of the model more scientifically, the average accuracy of 10 experiments conducted on each data set is calculated as the final criterion. In Figure 11, the OTGWO-WTELM showed the highest average accuracy value of 98.08% on the QSAR Biodegradation data set. The accuracy of MFO-WTELM and GWO-WTELM were 97.65% and 96.98%, respectively, respectively. In comparison, the accuracy of PSO-WTELM and WTELM was slightly lower, with 95.67% and 94.03%, respectively. On the other three data sets, the OTGWO-WTELM model also had the highest accuracy, reaching a maximum of 99.23%. Overall, the OTGWO-WTELM achieves the highest average accuracy results on different datasets. This indicates that the model has good generalization ability and stability. The GWO-WTELM and MFO-WTELM also show high accuracy on most data sets. However, WTELM performs relatively poorly on some datasets, confirming the necessity of using algorithms to optimize it.

To verify the classification stability of the model, the results obtained from 10 experiments on the data set are plotted as a box plot, as shown in Figure 12. The shorter block diagram in Figure 12 indicates that the model is more stable. A longer box indicates a greater fluctuation in model accuracy. From Figure 12, the OTGWO-WTELM

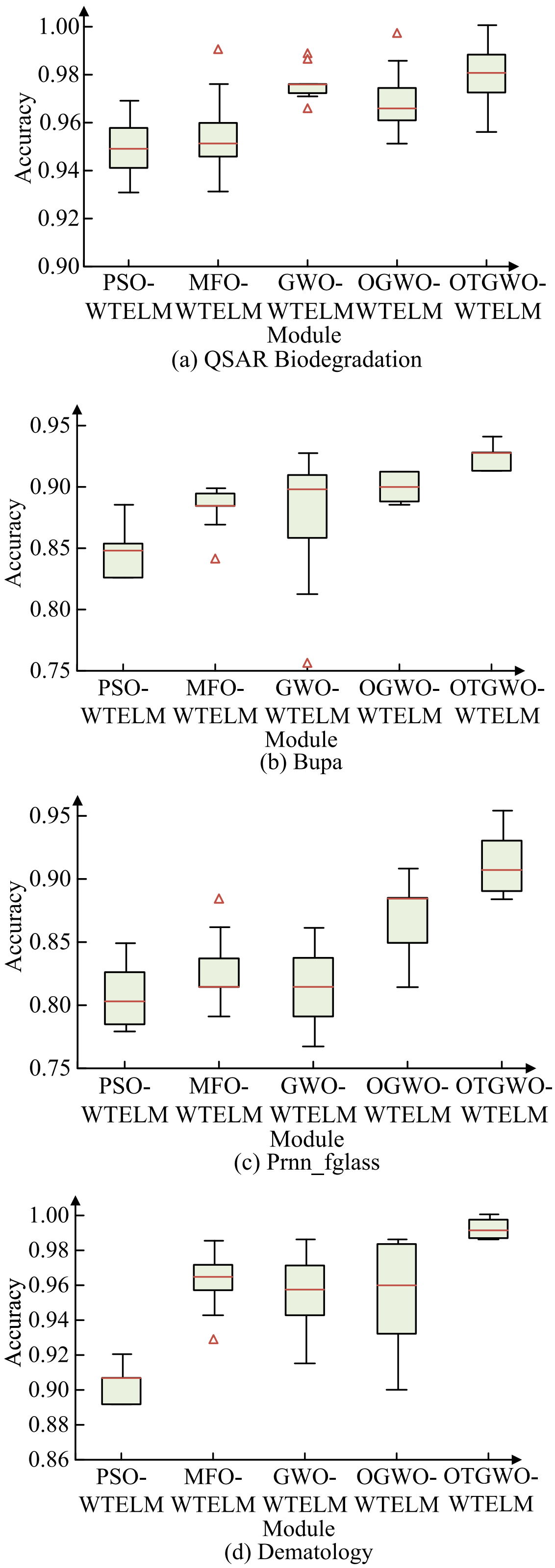

FIGURE 12. **Stability analysis of the model after algorithm optimization.**

model exhibited good stability on the Bupa data set and Dermatology data set, outperforming the other three models. It achieved the best results in stability. Although the stability of the OTGWO-WTELM model is not the best on the other two data sets, it is basically the same as the other models. The stability performance of GWO-WTELM model and OGWO-WTELM model is also relatively good. This indicates that the improved GWO can control the experimental results within a certain range during training the optimized model, despite there are randomly initialized parameters. At the same time, the improved GWO algorithm not only enhances the ability to optimize parameters, but also maintains the same or even better stability as other optimization algorithms.

In the experiment of exploring the convergence performance of the OTGWO-WTELM model, the maximum iteration of the population in the OTGWO algorithm is changed to 500. The maximum population is 100. The HL nodes in each layer of the WTELM model are adjusted to 100. The objective function of the optimization algorithm is set to the classification accuracy. Its purpose is to accurately transmit information to the optimization algorithm, so that it iteratively updates towards the goal. During this process, whether the objective function can converge to a stable value during model update iterations is an important basis for determining the convergence of the optimization algorithm. The convergence curves of different models are shown in Figure 13. From Figure 13, the OTGWO algorithm continuously updated the optimal parameters for WTELM with an increase in the number of iterations, allowing the fitness value to converge around a stable value. As shown in Figure 13 (a), in the QSAR Biodegradation data set, the OTGWO algorithm updated the optimal parameters of WTELM after approximately 180 iterations, and its fitness value increased to 0.868. When iterating to around 470 times, the fitness value increased to 0.931. According to Figure 13 (b), in the Bupa data set, when the OTGWO algorithm was iterated around 90 times, the fitness value increased to 0.951 and tended to stabilize. As shown in Figure 13 (c), in the Prnn_fglass data set, the fitness value of the OTGWO algorithm increased to 0.987 after approximately 50 iterations. As shown in Figure 13 (d), the OTGWO algorithm tended to stabilize after only 5 iterations, and its fitness value was infinitely close to 1. This indicates that the OTGWO algorithm has good convergence.

In addition, FMM-WOA-ELM-GMM [29], CS-ELM [30] and PCA-SDWPSO-ELM [31] are compare with the proposed IGWO-WTELM. All algorithms are trained using UCI data sets. Then, considering the over-fitting problem of the model, the model was tested using three independent data sets: Fashion-MNIST, CIFAR-10, and Iris Flower. Among them, Fashion-MNIST is a clothing classification data set containing 28×28 grayscale images. CIFAR-10 is an image classification data set with 10 categories. The Iris Flower data set is a plant classification data set that includes various categories. The test content is the accuracy and Root Mean Square Error (RMSE) of the algorithm. The test results are shown in Table 4. From Table 4, the accuracy of the

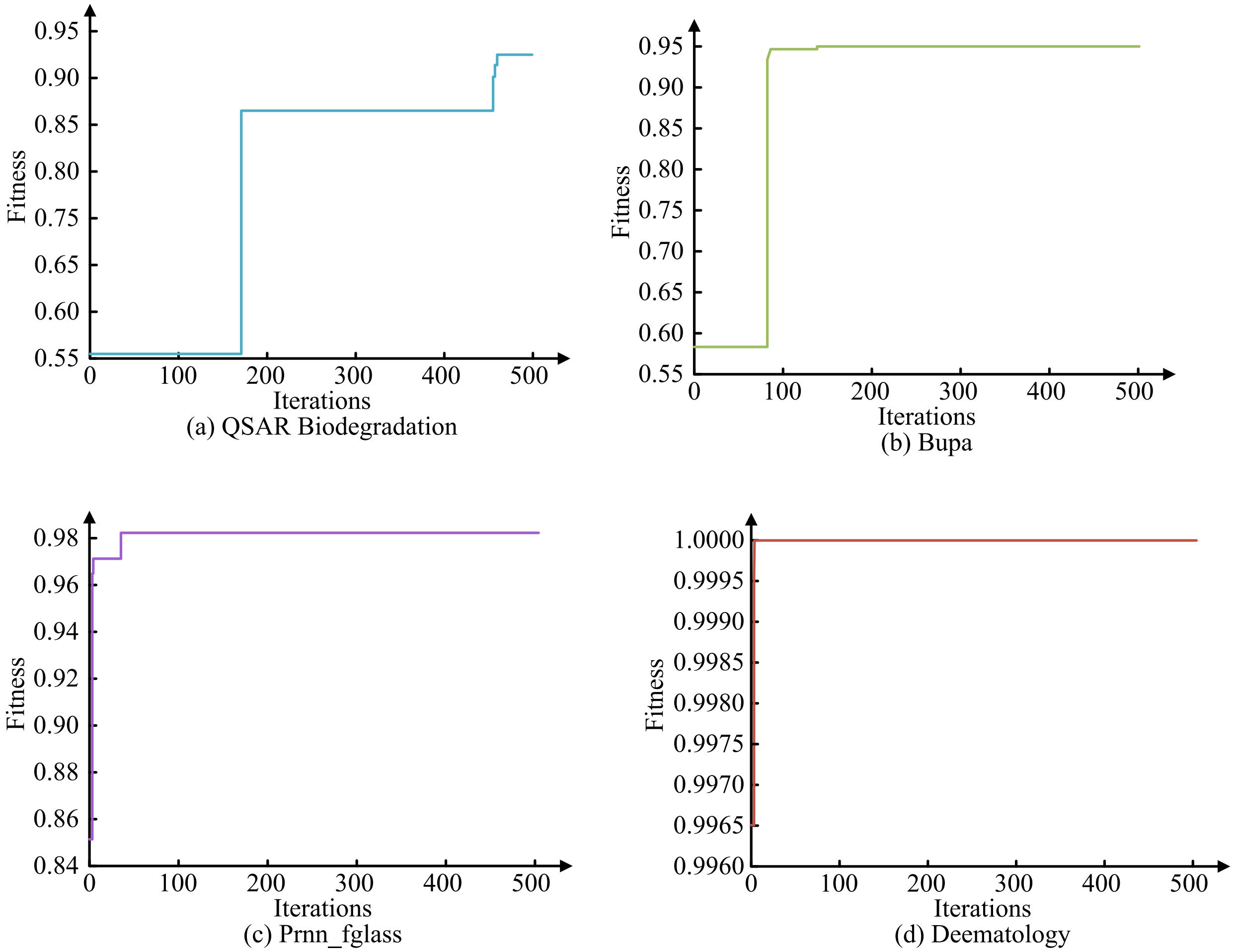

FIGURE 13. Convergence analysis of the model after algorithm optimization.

TABLE 5. Independent data set test results.

Algorithm index	OTGWO-WTELM	FCM–WOA–ELM–GMM	CS-ELM	PCA-SDWPSO-ELM
Year	2024	2023	2023	2023
Accuracy	The highest	Higher	Intermediate	Intermediate
Computational complexity	Moderation	Higher	Higher	Higher
Applicability	Best	Moderation	Moderation	Poor
Training and forecast time	Fastest	Faster	Intermediate	Intermediate

proposed IGWO-WTELM was better than the other three algorithms, leading by 7.2%-18.2%, and its RMSE value was only 0.021.

The operational efficiency of the model is measured by the training time. The original WTELM model does not require optimization iterations. The running time is much shorter than other models. Therefore, the running time of the WTELM model is not considered in the model efficiency. In this experiment, the maximum population of all models is 500. The maximum iterations are 100. The HL nodes in each layer are 100. The operational efficiency of each model is shown in Figure 14. From Figure 14 (a), on the QSAR Biodegradation data set, the GWO-WTELM model had the highest runtime, reaching 40.28%. The running time of the MFO-WTELM model and OTGWO-WTELM model was relatively low, at 13.88% and 23.09%, respectively. In Figure 14 (b), there was no significant difference in the proportion of running time between each model on the Bupa data set. The OTGWO-WTELM model had a running time ratio of 11.59%, with the best running efficiency. The

OTGWO-WTELM model in Prnn_fglass was also the lowest, only at 4.89%.

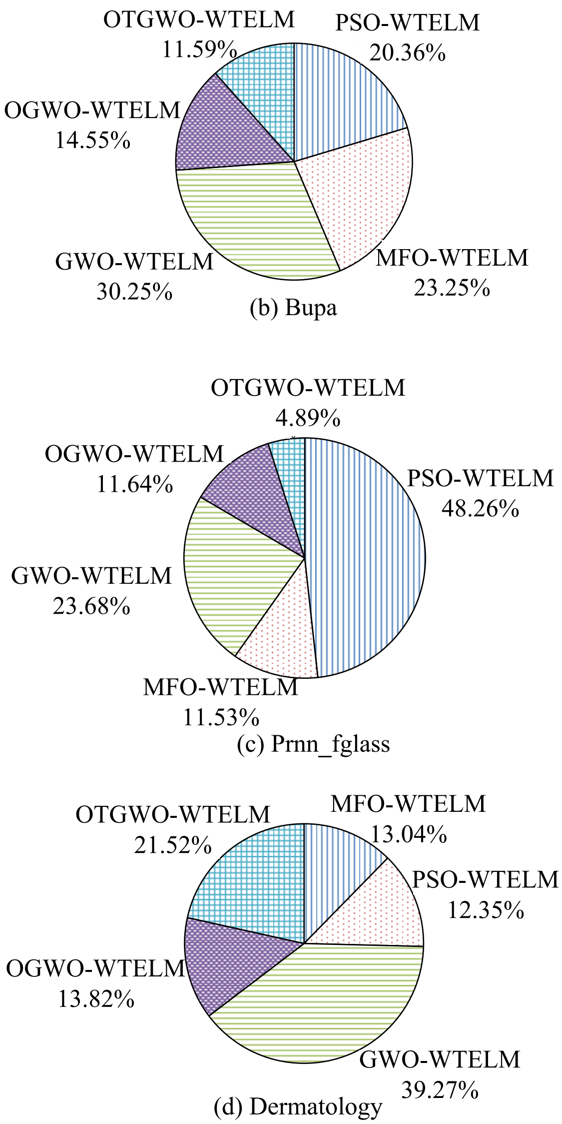

FIGURE 14. Analysis of the running efficiency of the model after algorithm optimization.

In Table 5, the comparison of several advanced methods is summarized. From Table 5, the method proposed in the study had the best performance. Based on the results obtained from various data sets, the OTGWO-WTELM model showed high operational efficiency. Especially in Prnn_fglass and Dermatology datasets, the running time of this model was significantly lower than that of other models. The experimental results show that the OTGWO-WTELM model has advantages such as high classification accuracy, relative stability, high operational efficiency, and fast convergence speed. This provides a new method for solving practical classification problems in the future, which has great application value.

IV. CONCLUSION

The research on classification issues has always received widespread attention. There are many models in machine learning methods that can effectively solve classification problems. The ELM has high classification accuracy. However, there are issues with parameter uncertainty and high demand for HL nodes. Therefore, based on the traditional ELM, the WTELM model was proposed to improve the classification accuracy by introducing the weighted factor. The GWO was enhanced by reverse learning and exploratory perception strategy. Then the improved algorithm was used to further optimize the model. The classification accuracy of the OTGWO-WTELM model in a single experiment reached 100%, the lowest value was 88.40%, and the average value was 95.26%. Moreover, in operational efficiency, the proposed model had a running time ratio of less than 25% on different datasets, with a minimum of 4.89%. The WTELM model proposed in the study outperformed ELM, TELM, and RELM in classification accuracy, stability, and running time. The intelligent optimization algorithm improved the classification accuracy of the model. Especially, the ability of the improved GWO to find the optimal solution was superior to other common algorithms. The improved OTGWO-WTELM has shown good performance in classification accuracy, stability, and operational efficiency, which has important application value. The parameter selection of the model may have a great influence on its performance. Parameter selection and adjustment are often based on trial and error or empirical judgment, which does not always guarantee an optimal model configuration. At the same time, the model may over-fit specific features of the training data, resulting in limited generalization ability. In addition, without special emphasis on feature selection and dimensionality reduction, the model may process some unnecessary information, which affects its performance. Future research can focus on developing methods that can automatically adjust parameters, exploring different weighted strategies to better handle data features, and applying feature selection and dimensionality reduction techniques to reduce model complexity. In addition, the model application in more complex classification tasks and the validation of generalization ability on different datasets are also potential research directions. These studies will help improve the classification accuracy, stability and operation efficiency of the model, making it more suitable for various practical application scenarios.

REFERENCES

[1] S. Sharma, K. Verma, and P. Hardaha, ‘‘Implementation of artificial intelligence in agriculture,’’ *J. Comput. Cognit. Eng.*, vol. 2, no. 2, pp. 155–162, Jul. 2022, doi: 10.47852/bonviewjcce2202174.

[2] M. Mirbabaie, S. Stieglitz, and N. R. J. Frick, ‘‘Artificial intelligence in disease diagnostics: A critical review and classification on the current state of research guiding future direction,’’ *Health Technol.*, vol. 11, no. 4, pp. 693–731, Jul. 2021, doi: 10.1007/s12553-021-00555-5.

[3] D. A. P. Pandian, ‘‘Artificial intelligence application in smart warehousing environment for automated logistics,’’ *December*, vol. 2019, no. 2, pp. 63–72, Dec. 2019, doi: 10.36548/jaicn.2019.2.002.

[4] I. Ibrahim and A. Abdulazeez, "The role of machine learning algorithms for diagnosing diseases," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 1, pp. 10–19, Mar. 2021, doi: 10.38094/jastt20179.

[5] M. A. Myszczynska, P. N. Ojamies, A. M. B. Lacoste, D. Neil, A. Saffari, R. Mead, G. M. Hautbergue, J. D. Holbrook, and L. Ferraiuolo, "Applications of machine learning to diagnosis and treatment of neurodegenerative diseases," *Nature Rev. Neurol.*, vol. 16, no. 8, pp. 440–456, Aug. 2020, doi: 10.1038/s41582-020-0377-8.

[6] P. Wang, E. Fan, and P. Wang, "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning," *Pattern Recognit. Lett.*, vol. 141, pp. 61–67, Jan. 2021, doi: 10.1016/j.patrec.2020.07.042.

[7] A. Pashaei, M. Ghatee, and H. Sajedi, "Convolution neural network joint with mixture of extreme learning machines for feature extraction and classification of accident images," *J. Real-Time Image Process.*, vol. 17, no. 4, pp. 1051–1066, Aug. 2020, doi: 10.1007/s11554-019-00852-3.

[8] H. Zhao, H. Liu, J. Xu, and W. Deng, "Performance prediction using high-order differential mathematical morphology gradient spectrum entropy and extreme learning machine," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4165–4172, Jul. 2020, doi: 10.1109/TIM.2019.2948414.

[9] J. Li and Y. Wu, "Improved sparrow search algorithm with the extreme learning machine and its application for prediction," *Neural Process. Lett.*, vol. 54, no. 5, pp. 4189–4209, Oct. 2022, doi: 10.1007/s11063-022-10804-x.

[10] A. Gumaei, M. M. Hassan, M. R. Hassan, A. Alelaiwi, and G. Fortino, "A hybrid feature extraction method with regularized extreme learning machine for brain tumor classification," *IEEE Access*, vol. 7, pp. 36266–36273, 2019, doi: 10.1109/ACCESS.2019.2904145.

[11] X. Zhao, M. Jia, P. Ding, C. Yang, D. She, and Z. Liu, "Intelligent fault diagnosis of multichannel motor–rotor system based on multimanifold deep extreme learning machine," *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 5, pp. 2177–2187, Oct. 2020, doi: 10.1109/TMECH.2020.3004589.

[12] F. T. Al-Dhief, M. M. Baki, N. M. A. Latiff, N. N. N. A. Malik, N. S. Salim, M. A. A. Albader, N. M. Mahyuddin, and M. A. Mohammed, "Voice pathology detection and classification by adopting online sequential extreme learning machine," *IEEE Access*, vol. 9, pp. 77293–77306, 2021, doi: 10.1109/ACCESS.2021.3082565.

[13] N. Li, F. He, W. Ma, R. Wang, and X. Zhang, "Wind power prediction of kernel extreme learning machine based on differential evolution algorithm and cross validation algorithm," *IEEE Access*, vol. 8, pp. 68874–68882, 2020, doi: 10.1109/ACCESS.2020.2985381.

[14] F. Zhang, S. Tian, S. Chen, Y. Ma, X. Li, and X. Guo, "Voxel-based morphometry: Improving the diagnosis of Alzheimer's disease based on an extreme learning machine method from the ADNI cohort," *Neuroscience*, vol. 414, pp. 273–279, Aug. 2019, doi: 10.1016/j.neuroscience.2019.05.014.

[15] M. Shariati, M. S. Mafipour, B. Ghahremani, F. Azarhomayun, M. Ahmadi, N. T. Trung, and A. Shariati, "A novel hybrid extreme learning machine–grey wolf optimizer (ELM-GWO) model to predict compressive strength of concrete with partial replacements for cement," *Eng. With Comput.*, vol. 38, no. 1, pp. 757–779, Feb. 2022, doi: 10.1007/s00366-020-01081-0.

[16] M. A. Salam, A. T. Azar, and R. Hussien, "Swarm-based extreme learning machine models for global optimization," *Comput., Mater. Continua*, vol. 70, no. 3, pp. 6339–6363, Jul. 2022, doi: 10.32604/cmc.2022.020583.

[17] C. Li, J. Zhou, M. Khandelwal, X. Zhang, M. Monjezi, and Y. Qiu, "Six novel hybrid extreme learning machine–swarm intelligence optimization (ELM–SIO) models for predicting backbreak in open-pit blasting," *Natural Resour. Res.*, vol. 31, no. 5, pp. 3017–3039, Jun. 2022, doi: 10.1007/s11053-022-10082-3.

[18] K. Roushangar, A. A. Sadaghiani, and S. Shahnazi, "Novel application of robust GWO-KELM model in predicting discharge coefficient of radial gates: A field data-based analysis," *J. Hydroinformatics*, vol. 25, no. 2, pp. 275–299, Feb. 2023, doi: 10.2166/hydro.2023.096.

[19] A. Diker, Y. Sönmez, F. Özyurt, E. Avcı, and D. Avcı, "Examination of the ECG signal classification technique DEA-ELM using deep convolutional neural network features," *Multimedia Tools Appl.*, vol. 80, no. 16, pp. 24777–24800, Apr. 2021, doi: 10.1007/s11042-021-10517-8.

[20] K. Jain and A. Saxena, "Simulation on supplier side bidding strategy at day-ahead electricity market using ant lion optimizer," *J. Comput. Cognit. Eng.*, vol. 2, no. 1, pp. 17–27, Mar. 2022, doi: 10.47852/bonviewjcce2202160.

[21] W. Liu, Z. Wang, Y. Yuan, N. Zeng, K. Hone, and X. Liu, "A novel sigmoid-function-based adaptive weighted particle swarm optimizer," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 1085–1093, Feb. 2021, doi: 10.1109/TCYB.2019.2925015.

[22] X.-F. Song, Y. Zhang, D.-W. Gong, and X.-Z. Gao, "A fast hybrid feature selection based on correlation-guided clustering and particle swarm optimization for high-dimensional data," *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 9573–9586, Sep. 2022, doi: 10.1109/TCYB.2021.3061152.

[23] Y. Tikhamarine, D. Souag-Gamane, A. N. Ahmed, O. Kisi, and A. El-Shafie, "Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey wolf optimization (GWO) algorithm," *J. Hydrol.*, vol. 582, Mar. 2020, Art. no. 124435, doi: 10.1016/j.jhydrol.2019.124435.

[24] A. H. Ali, M. G. Yaseen, M. Aljanabi, S. A. Abed, and C. Gpt, "Transfer learning: A new promising techniques," *Mesopotamian J. Big Data*, vol. 2023, pp. 29–30, Feb. 2023, doi: 10.58496/mjbd/2023/004.

[25] R. Rajamoorthy, G. Arunachalam, P. Kasinathan, R. Devendiran, P. Ahmadi, S. Pandiyan, S. Muthusamy, H. Panchal, H. A. Kazem, and P. Sharma, "A novel intelligent transport system charging scheduling for electric vehicles using grey wolf optimizer and sail fish optimization algorithms," *Energy Sources, A, Recovery, Utilization, Environ. Effects*, vol. 44, no. 2, pp. 3555–3575, Jun. 2022, doi: 10.1080/15567036.2022.2067268.

[26] A. K. Sales, E. Gul, M. J. S. Safari, H. G. Gharehbagh, and B. Vaheddoost, "Urmia lake water depth modeling using extreme learning machine-improved grey wolf optimizer hybrid algorithm," *Theor. Appl. Climatol.*, vol. 146, nos. 1–2, pp. 833–849, Oct. 2021, doi: 10.1007/s00704-021-03771-1.

[27] M. O. F. Goni, M. N. I. Mondal, S. M. R. Islam, M. Nahiduzzaman, M. R. Islam, M. S. Anower, and K.-S. Kwak, "Diagnosis of malaria using double hidden layer extreme learning machine algorithm with CNN feature extraction and parasite inflator," *IEEE Access*, vol. 11, pp. 4117–4130, 2023, doi: 10.1109/ACCESS.2023.3234279.

[28] Y.-L. He, L. Chen, Y. Gao, J.-H. Ma, Y. Xu, and Q.-X. Zhu, "Novel double-layer bidirectional LSTM network with improved attention mechanism for predicting energy consumption," *ISA Trans.*, vol. 127, pp. 350–360, Aug. 2022, doi: 10.1016/j.isatra.2021.08.030.

[29] B. Gu, H. Hu, J. Zhao, H. Zhang, and X. Liu, "Short-term wind power forecasting and uncertainty analysis based on FCM–WOA–ELM–GMM," *Energy Rep.*, vol. 9, pp. 807–819, Dec. 2023, doi: 10.1016/j.egyr.2022.11.202.

[30] Z. Wang, N. Wang, H. Zhang, L. Jia, Y. Qin, Y. Zuo, Y. Zhang, and H. Dong, "Segmentalized mRMR features and cost-sensitive ELM with fixed inputs for fault diagnosis of high-speed railway turnouts," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 5, pp. 4975–4987, May 2023, doi: 10.1109/TITS.2023.3239636.

[31] Z. Wang, Y. Liu, T. Wang, D. Gong, and D. Zhang, "Prediction model of hot strip crown based on industrial data and hybrid the PCA-SDWPSO-ELM approach," *Soft Comput.*, vol. 27, no. 17, pp. 12483–12499, Feb. 2023, doi: 10.1007/s00500-023-07895-6.

NA ZHANG was born in Henan, China, in 1980. She received the Ph.D. degree from Donghua University, Shanghai, China, in 2014.

From 2007 to 2020, she taught with Henan College of Urban Construction, China. Since 2021, she has been an Associate Professor with the School of Computer and Artificial Intelligence, Henan Finance University. She has published more than ten articles as the first author. She presided more than the completed five provincial projects. She completed two monographs. Her research interests include machine learning, deep learning, and natural language processing.

XIAOFENG WANG was born in Henan, China, in 1980. He received the M.S. degree in applied architecture and civil engineering from Henan Polytechnic University, Henan, in 2019. He is currently pursuing the Ph.D. degree with SEGi University, Malaysia.