



Selective ensemble of doubly weighted fuzzy extreme learning machine for tumor classification

Yang Wang¹

Received: 11 March 2023 / Accepted: 25 March 2024 / Published online: 8 May 2024
© Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Malignant epithelial cell tumor also known as cancer is a deadly disease requiring a very costly and complex treatment. Early and accurate diagnosis of tumor plays an important role in reducing the mortality rate. With the rapid development of gene chip technology, gene expression data based tumor classification is helpful for accurate decision-making and has achieved great attention of researchers. Due to gene expression data having the properties of multi-class imbalance, high noise and high-dimensional small samples, in this paper, selective ensemble of doubly weighted fuzzy extreme learning machine (SEN-DWFELM) is presented for tumor classification. In view of good generalization performance of extreme learning machine (ELM), feature weighted fuzzy membership is embedded in ELM for eliminating classification error from noise samples. It considers the influence of feature importance on classification to acquire more accurate fuzzy membership. Simultaneously, by removing features with smaller weights it reduces the dimensionality of samples to improve training efficiency. Considering imbalanced learning, the weighted scheme is also introduced to enhance the effect of minority class samples on classification. Furthermore, doubly weighted fuzzy extreme learning machine (DWFELM) based selective ensemble algorithm is proposed to make classification performance more robust. Partial-based DWFELMs are selected using binary version of an improved whale optimization algorithm, and the selected base DWFELMs are integrated by majority voting. Finally, the proposed SEN-DWFELM is compared with conventional ensemble methods and variants of SEN-DWFELM on various gene expression data. Experimental results show that SEN-DWFELM remarkably outperforms other competitors in accordance with classification performance and can effectively deal with tumor diagnosis problems.

Keywords Gene expression data · Feature weighted · Fuzzy extreme learning machine · Selective ensemble · Whale optimization algorithm

1 Introduction

Tumor includes malignant and benign tumor, where malignant epithelial cell tumor also known as cancer is a hazard to human health [1]. So accurate and timely treatment of cancer is of vital importance in reducing the mortality rate. Nevertheless, the same cancer caused by many factors can have different symptoms, and it is difficult for traditional methods to identify cancer precisely [2]. With the rapid development of gene chip technology, gene expression data based tumor classification can acquire more accurate results and has drawn a lot of research interests [3]. In virtue of gene

expression data having the properties of multi-class imbalance, high noise and high-dimensional small samples, it is indispensable to carry out research on gene expression classification.

Extreme learning machine (ELM) proposed by Huang et al. [4] is a new type of single hidden layer feedforward neural networks (SLFNs). In virtue of good generalization performance with a fast learning speed, ELM has been successfully applied to various practical application [5–8]. Inspired by the ensemble idea [9, 10], the stability along with generalization performance of single ELM can be further improved. AdaBoost constructed a serial ensemble model. The weight distributed for every training sample is updated in the light of classification performance of the former classifiers before every iteration [11]. Li et al. [12] proposed that weighted ELM is put into the amended AdaBoost framework, namely boosting weighted ELM. The weight distributed for every

✉ Yang Wang
wangyang0531@163.com

¹ School of Artificial Intelligence and Software, Liaoning Petrochemical University, Fushun 113001, China

training sample from different class is updated apart. Cao et al. [13] proposed V-ELM that completes ensemble of ELMs and then reaches the decision by majority voting. Lu et al. [14] proposed DF-D-ELM and D-D-ELM that select some base ELMs by the dissimilarity measure and then integrate the selected ELMs by majority voting.

Nowadays, gene expression classification brings extreme challenge. Existing methods usually ignore the influence of noises and outliers on classification and are extremely sensitive to noises and outliers. Zhang and Ji [15] introduced a fuzzy membership to solve the problem. The fuzzy membership can be flexibly set on different applications. Moreover, existing methods usually assume balanced class distribution and are unfit for handling data with complex class distribution. Data processing approach and algorithmic approach are two strategies handling imbalanced dataset [16]. In data processing approach, there is mainly under-sampling and oversampling studied for balancing the size of every class [17, 18]. Data processing approach changes samples distribution and probably brings about losing crucial information, whereas algorithmic approach has come into wide use in handling imbalanced dataset without changing samples distribution [19, 20]. Gupta [21] proposed weighted twin support vector regression based on K-nearest neighbor (KNN), which decreases the effect of outliers and achieves better generalization performance. Hazarika et al. [22] proposed density-weighted twin support vector machine (SVM) considering 2-norm of slack variables and equality constraints, which validate the usability and efficacy of the proposed model for binary class imbalance learning.

This paper takes an interest in algorithmic approach, and presents selective ensemble of doubly weighted fuzzy extreme learning machine (SEN-DWFELM) for tumor classification. Firstly, doubly weighted fuzzy extreme learning machine (DWFELM) is constructed. ReliefF is a high-efficiency feature optimization algorithm, which can effectively cope with noisy data and missing data without limiting data types [23]. ReliefF-based feature weighted fuzzy membership is developed to eliminate classification error from noises and outliers. It not only considers the decrease in the dimensionality of samples but also considers the impact of feature importance on classification. Simultaneously, the weighted scheme is introduced to enhance the relative influence of minority class samples and lessen the bias against performance from imbalanced dataset.

Furthermore, binary version of an improved whale optimization algorithm (IWOA) is put forward for selecting some base DWFELMs integrated by majority voting. Whale optimization algorithm (WOA) is a new meta-heuristic method imitating hunting behaviors of humpback whales [24]. With good search capability and simplicity property, WOA has been widely used for tackling various optimization problems [25–29]. However, the property of WOA has still much

room for improving as described by Mafarja et al. [30]. Therefore, in this paper, population initialization based on quasi-opposition learning strategy is adopted to accelerate convergence. The dynamic weight and nonlinear control parameter are proposed for coordinating the exploitation and exploration ability. By various experiments, comparison results show the proposed SEN-DWFELM can acquire better classification performance and is suited for coping with gene expression data.

The remaining section of this paper is arranged below. Section 2 discusses related works based on gene expression classification. Section 3 presents the description of preliminaries. In Sect. 4, the proposed method is explained in detail. In Sect. 5, the experimental design and comparative results of related algorithms are presented. Finally, Sect. 6 states conclusions of the paper.

2 Related works

Recently, many methods have been developed for tackling gene expression classification problems. Gao et al. [31] presented a hybrid gene selection method, where information gain (IG) is initially used for filtering redundant and irrelevant genes, then SVM is used for further removing redundant genes and eliminating noises. Selected genes are again served as input for SVM classifier, and the proposed method achieves superior classification performance. Rani et al. [32] presented a two-stage gene selection method, where mutual information (MI) is firstly employed for selecting genes having higher information relevant to cancers, then genetic algorithm (GA) is again employed for identifying the optimal set of genes in the second stage. Besides, SVM is employed for classification and the proposed method acquires higher classification accuracy. Tavasoli et al. [33] introduced an optimized gene selection method, where Wilcoxon, two-sample T-test, entropy, receiver operating characteristic curve, and Bhattacharyya distance are employed in the ensemble soft weighted approach. Moreover, parameter optimization of SVM is performed by a modified water cycle algorithm and the proposed method shows its robustness in terms of accuracy. Lu et al. [34] presented an efficient gene selection method combining the adaptive GA and the mutual information maximization (MIM), which remarkably removed the redundancies of genes. Four different classifiers are used for classification, then the proposed method shows reduced genes provide higher classification accuracy. Mondal et al. [35] used entropy-based method for differentiating between breast tumor and normal tissue. In addition, random forest, naive Bayes, KNN and SVM are used for breast cancer prediction. Experimental results show SVM obtains better accuracy. Shukla et al. [36] introduced a new wrapper gene selection method, where minimum redundancy maximum

relevance (mRMR) is firstly used for selecting relevant genes from gene expression data, then teaching learning-based algorithm in combination with gravitational search algorithm is employed to select the informative genes from data reduced by mRMR. Naive Bayes classifier is again employed to classify cancer, and experimental results demonstrate its effectiveness regarding optimal number of genes and classification accuracy. Dabba et al. [37] incorporated modified moth flame algorithm in MIM to evolve gene subsets and used SVM for detecting cancer. The proposed method provides greater classification accuracy. All these algorithms have focused on solving gene selection problem. However, confronted with the properties of multi-class imbalance, high noise and high-dimensional small samples of gene expression data, it is crucial to find an appropriate classifier for addressing these problems.

3 Preliminaries

3.1 Weighted extreme learning machine (WELM)

Given a training dataset comprising N different samples (x_j, z_j) , where $x_j = [x_{j1}, x_{j2}, \dots, x_{jm}]^T \in R^m$ is an $m \times 1$ feature vector and $z_j = [z_{j1}, z_{j2}, \dots, z_{jn}]^T \in R^n$ is an $n \times 1$ target vector. With activation function $G(x)$ and hidden nodes L , the mathematical pattern of SLFNs can be expressed as follows.

$$\sum_{k=1}^L \beta_k G(a_k \cdot x_j + b_k) = z_j \quad j = 1, 2, \dots, N \quad (1)$$

where $a_k = [a_{k1}, a_{k2}, \dots, a_{km}]^T$ means the input weight vector for the link between the k th hidden node and input nodes, $a_k \cdot x_j$ means the inner product for a_k and x_j , b_k means the bias for the k th hidden node, and $\beta_k = [\beta_{k1}, \beta_{k2}, \dots, \beta_{kn}]^T$ means the output weight vector for the link between output nodes and the k th hidden node.

Training samples can be approximated with zero error in SLFNs if L is identical with N . Then the formula (1) can also be given as follows.

$$H\beta = Z \quad (2)$$

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} G(a_1 \cdot x_1 + b_1) & \dots & G(a_L \cdot x_1 + b_L) \\ \vdots & \dots & \vdots \\ G(a_1 \cdot x_N + b_1) & \dots & G(a_L \cdot x_N + b_L) \end{bmatrix}_{N \times L}$$

Given a training dataset $(x_j, z_j), j = 1, 2, \dots, N$, hidden nodes L and activation function $G(x)$.

Step 1. Generate the hidden layer biases b_k and input weights a_k randomly, $k = 1, 2, \dots, L$.

Step 2. Calculate the output matrix for hidden layer H .

Step 3. Calculate the output weight $\beta = H^+Z$.

Fig. 1 Algorithm for ELM

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times n}, \text{ and } Z = \begin{bmatrix} z_1^T \\ \vdots \\ z_N^T \end{bmatrix}_{N \times n} \quad (3)$$

where β means the output weight matrix, and Z means the output matrix. H means the output matrix for hidden layer, and the k th column for H indicates the output vector for the k th hidden node concerning all of the inputs.

Nevertheless, Eq. (2) can not be fulfilled owing to $L < N$ in most cases. The output weights are just calculated through the Least Square solution $\beta = H^+Z$ using the Moore-Penrose generalized inverse H^+ of H [38, 39]. ELM algorithm is described briefly in Fig. 1.

In light of Bartlett’s theory [40], the aim of ELM is to minimize training errors and the norm of output weights. Then an especial weight is also assigned for every sample to better handle imbalanced dataset. Consequently, the classification problem of weighted extreme learning machine (WELM) is defined below.

$$\min : L = \frac{1}{2} \|\beta\|^2 + CW \frac{1}{2} \sum_{j=1}^N \xi_j^2 \quad (4)$$

$$\text{s.t : } h(x_j)\beta = z_j - \xi_j$$

where C is penalty parameter, and $\xi_j = [\xi_{j1}, \xi_{j2}, \dots, \xi_{jn}]^T$ is the training error vector of all the output nodes on the training sample x_j . As a diagonal matrix, W is associated with every training sample and can be assigned below [41].

$$W1 : w_{jj} = \frac{1}{\#(z_j)} \quad (5)$$

$$W2 : w_{jj} = \begin{cases} \frac{0.618}{\#(z_j)} & \text{if } \#(z_j) > \text{avg} \\ \frac{1}{\#(z_j)} & \text{if } \#(z_j) \leq \text{avg} \end{cases}$$

where avg is the average of the samples number of all the classes, and $\#(z_j)$ is the samples number of class z_j . Based on KKT theorem, the output weight of WELM is calculated as

$$\beta = \begin{cases} H^T (\frac{1}{C} + WHH^T)^{-1} WZ & \text{when } N < L \\ (\frac{1}{C} + H^T WH)^{-1} H^T WZ & \text{when } N \gg L \end{cases} \quad (6)$$

3.2 Whale optimization algorithm (WOA)

WOA simulates hunting behaviors of humpback whales, namely bubble-net feeding approach. There are mainly three stages including searching for prey, encircling prey and bubble-net attacking in WOA [24]. The mathematical model is described as follows.

3.2.1 Searching for prey

In the exploration stage of searching for prey, whales seek randomly according to respective position to perform a global search. This behavior is written as

$$\vec{D} = |\vec{F} \cdot \vec{U}_{\text{rand}} - \vec{U}(t)| \quad (7)$$

$$\vec{U}(t+1) = \vec{U}_{\text{rand}} - \vec{A} \cdot \vec{D} \quad (8)$$

where \vec{U}_{rand} is a whale position of random selection from the present population, \vec{U} is the present position for a whale, t is the present iteration, \vec{D} is the distance kept by the random position and the present position, then the coefficient vectors \vec{F} and \vec{A} are defined as

$$\vec{F} = 2 \cdot \vec{r} \quad (9)$$

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \quad (10)$$

where \vec{r} is a random vector in $[0,1]$, $\vec{a} = 2 - t * (2/MN)$ is decreasing linearly from 2 to 0, and MN is the maximum iteration number.

3.2.2 Encircling prey

In the exploitation stage of encircling prey, the target prey is just the present best solution, which offers guidance on other solutions updating their positions. This behavior is expressed as

$$\vec{D} = |\vec{F} \cdot \vec{U}_{\text{best}} - \vec{U}(t)| \quad (11)$$

$$\vec{U}(t+1) = \vec{U}_{\text{best}} - \vec{A} \cdot \vec{D} \quad (12)$$

where \vec{U}_{best} is the present best solution. If a better solution is acquired after every iteration, \vec{U}_{best} is also updated accordingly.

3.2.3 Bubble-net attacking

Bubble-net attacking approach performs the entire exploitation stage. In this stage, WOA uses a spiral-shaped path of bubble nets to simulate attacking behaviors of humpback whales, and this behavior is shown below.

$$\vec{D}' = |\vec{U}_{\text{best}} - \vec{U}(t)| \quad (13)$$

$$\vec{U}(t+1) = \vec{D}' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{U}_{\text{best}} \quad (14)$$

where \vec{D}' is the distance kept by the prey and the whale, l is a random number in $[-1, 1]$, and b is a constant indicating the logarithmic spiral shape.

WOA starts with a group of random solutions. $|A|$ determines whether WOA emphasizes the exploitation or exploration ability. Whales update their positions by Eq. (8) if $|A| \geq 1$, and the exploration capability is emphasized. Whales update their positions relying on p by Eqs. (12) or (14) if $|A| < 1$, and the exploitation capability is emphasized. It is assumed that there is 50% probability to move between bubble-net attacking and encircling prey behavior. So this behavior is formulated as

$$\vec{U}(t+1) = \begin{cases} \vec{U}_{\text{best}} - \vec{A} \cdot \vec{D} & \text{if } p < 0.5 \\ \vec{D}' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{U}_{\text{best}} & \text{if } p \geq 0.5 \end{cases} \quad (15)$$

where p is a random number in $[0, 1]$. The flowchart of WOA is shown in Fig. 2.

4 Proposed SEN-DWFELM model

4.1 ReliefF algorithm

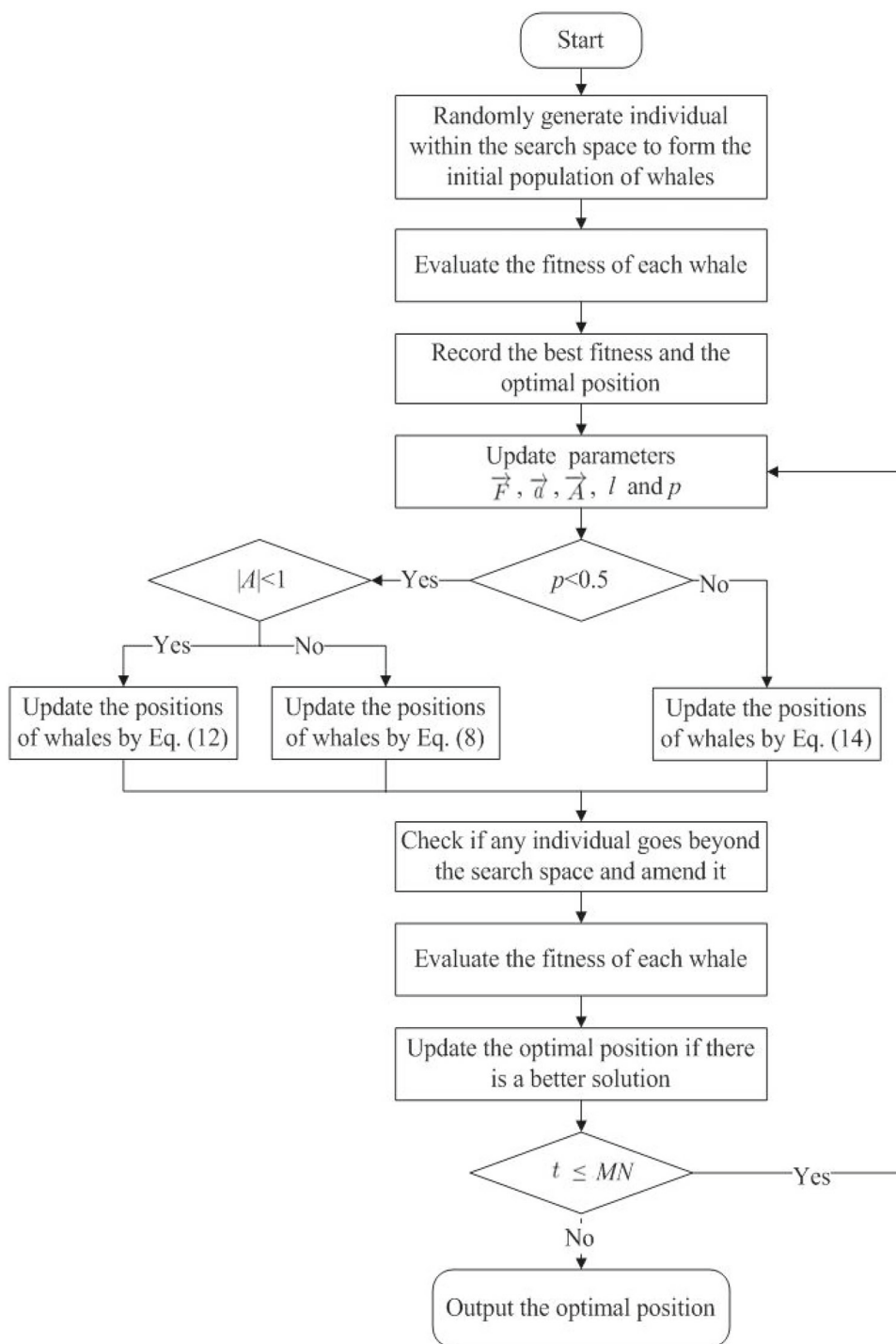
Recently, ReliefF algorithm is widely applied in feature optimization [42, 43]. It selects randomly a sample x_i , then searches for k -nearest neighbor samples of x_i from the same class called H_j and the different classes called $M_j(e)$. The initial weights assigned for all the features are 0. These weights are again calculated by the between-class and within-class distances kept by the nearest neighbor samples. The formula of adjusting the weights is expressed as

$$f_v = f_v - \sum_{j=1}^k \frac{\text{diff}(v, x_i, H_j)}{q \cdot k} + \sum_{e \neq \text{class}(x_i)} \frac{P(e)}{1 - P(\text{class}(x_i))} \cdot \sum_{j=1}^k \frac{\text{diff}(v, x_i, M_j(e))}{q \cdot k} \quad (16)$$

where $P(e)$ is the ratio of samples in the e th class to the total samples, $P(\text{class}(x_i))$ is the ratio of samples in the same class as x_i to the total samples, $\text{diff}(v, x_i, H_j)$ is the distance between x_i and H_j on the v th feature, $\text{diff}(v, x_i, M_j(e))$ is the distance between x_i and $M_j(e)$ on the v th feature, then q and k are respectively the amount of the sampling and the nearest neighbor samples.

ReliefF algorithm aims at repeating q times of the above procedure. Finally, the weights of all the features are cal-

Fig. 2 The flowchart of WOA



culated, and then feature optimization is completed by the features with higher weights.

4.2 Doubly weighted fuzzy extreme learning machine (DWFELM)

Feature weighted fuzzy membership and the weighted scheme are introduced into ELM in this paper. Feature

weighted fuzzy membership is proposed for eliminating classification error from noises and outliers. In [15], all the features are used for calculating fuzzy membership and are regarded as having the same contribution to classification. However, this method can assign high fuzzy membership for noises and outliers to result in reducing classification performance. Therefore, based on the feature set obtained by ReliefF algorithm, the weighted distance between training

sample and class center is developed for analyzing the impact of feature importance to improve classification performance.

First, the samples center for every class is expressed as

$$d_t^v = \frac{1}{N_t} \sum_{j=1}^{N_t} x_j^v \quad t = 1, \dots, c \tag{17}$$

where x_j^v is the v th feature of the j th sample, c is the classes number, N_t is the samples number for the t th class, and d_t^v is the samples center for the t th class. Feature weighted distance $f d_t$ for the t th class from samples to center is expressed below.

$$f d_t = \sqrt{\sum_{v=1}^m f_v (x_j^v - d_t^v)^2} \quad j = 1, \dots, N_t \tag{18}$$

where f_v is the weight of the v th feature, and m is the amount of feature. Accordingly, feature weighted fuzzy membership is presented below.

$$R : r_{jj} = 1 - \frac{f d_t}{\max(f d_t) + \epsilon} \tag{19}$$

where ϵ is very small and positive. As a diagonal matrix, R is closely associated with every training sample. From Eqs. (18) and (19), it can be found that feature weighted fuzzy membership avoids being dominated by some uncorrelated or weakly correlated features. In order to reduce the effect of noises and outliers, the approach will exert a minimum fuzzy membership over them.

On the other hand, the weighted scheme is designed for enhancing the relative influence of minority class samples, and it is presented as

$$W : w_{jj} = \frac{\#(c - z_j + 1)}{\max(\#z_j)} \tag{20}$$

where $\max(\#z_j)$ is the maximum samples number for all the classes, and $\#(c - z_j + 1)$ is the samples number for class $c - z_j + 1$. The samples number for class 1,2,..,c is arranged in the order of rising.

In contrast, W -based weighted scheme has much stronger influence on classification performance, and the reason is explained below for binary classification problem.

$$\begin{aligned} \Delta W &= \left(\frac{\#majority}{\max(\#z_j)} \right) - \left(\frac{\#minority}{\max(\#z_j)} \right) \\ &= \frac{\#majority - \#minority}{\max(\#z_j)} \\ \Delta W1 &= \left(\frac{1}{\#minority} \right) - \left(\frac{1}{\#majority} \right) \end{aligned}$$

$$= \frac{\#majority - \#minority}{\#majority \times \#minority}$$

where $\#majority$ and $\#minority$ denote respectively the amount of majority class and minority class samples. Due to $\#majority > 2$ and $\#minority > 2$, $\#majority \times \#minority > \max(\#z_j)$. So $\Delta W > \Delta W1$ and W -based weighted scheme is superior to $W1$ -based weighted scheme.

$$\begin{aligned} \Delta W2 &= \left(\frac{1}{\#minority} \right) - \left(\frac{0.618}{\#majority} \right) \\ &= \frac{\#majority - 0.618 \times \#minority}{\#majority \times \#minority} \end{aligned}$$

Due to $\#majority > 2$ and $\#minority > 2$, $\#majority \times \#minority - \max(\#z_j) > (\#majority - 0.618 \times \#minority) - (\#majority - \#minority)$. So $\Delta W > \Delta W2$ and W -based weighted scheme is superior to $W2$ -based weighted scheme. Consequently, the classification problem of DWFELM is expressed below.

$$L_{DWFELM} = \frac{1}{2} \|\beta\|^2 + CW \frac{1}{2} \sum_{j=1}^N r_{jj} \xi_j^2 \tag{21}$$

$$\text{s.t : } h(x_j)\beta = z_j - \xi_j$$

Relative constraint conditions are deduced below on the grounds of KKT theorem.

$$\frac{\partial L_{DWFELM}}{\partial \beta} = 0 \rightarrow \beta = \sum_{j=1}^N \alpha_j h(x_j)^T = H^T \alpha \tag{22}$$

$$\frac{\partial L_{DWFELM}}{\partial \xi_j} = 0 \rightarrow \alpha_j = CW r_{jj} \xi_j \tag{23}$$

$$\frac{\partial L_{DWFELM}}{\partial \alpha_j} = 0 \rightarrow h(x_j)\beta - z_j + \xi_j = 0 \tag{24}$$

If $N < L$, by substituting Eqs. (23) and (22) into (24), the output weight of DWFELM is calculated as

$$\beta = H^T \left(\frac{R^{-1}}{C} + WHH^T \right)^{-1} WZ \tag{25}$$

If $N \gg L$, by substituting Eqs. (23) and (24) into (22), the output weight of DWFELM is calculated as

$$\beta = \left(\frac{I}{C} + H^T WRH \right)^{-1} H^T WRZ \tag{26}$$

4.3 Improved Whale optimization algorithm (IWOA)

WOA has the risk of poor convergence because of diverse exploration and less exploitation [24]. In this paper, it is indispensable for searching for a well improved WOA (IWOA).

4.3.1 Population initialization based on quasi-opposition learning strategy

WOA does not guarantee the diversity of initialization population because it uses the approach of random initialization. Opposition-based learning (OBL) adopts the notion of opposite point [44], and then adds opposite search to random search for accelerating search speed. In general, OBL considers both initial solutions and opposite solutions to find the best solution faster.

In initialization phase, IWOA generates an initial population randomly, and every solution $U_j = \{u_{j,1}, u_{j,2}, \dots, u_{j,D}\}$ is expressed as

$$u_{j,k} = u_{\min,k} + \text{rand}(0, 1)(u_{\max,k} - u_{\min,k}) \tag{27}$$

$$j = 1, 2, \dots, PN \quad k = 1, 2, \dots, D$$

where $u_{\min,k}$ and $u_{\max,k}$ are lower and upper bound for the k th parameter, D is the amount of optimization parameters, and PN is the size of population. Then the opposite solution $OU_j = \{ou_{j,1}, ou_{j,2}, \dots, ou_{j,D}\}$ of U_j is expressed as

$$ou_{j,k} = u_{\min,k} + u_{\max,k} - u_{j,k} \tag{28}$$

with the research going on, Rahnamayan et al. [45] found a new method called quasi-opposition-based learning (QOBL). Compared with OBL, solutions of QOBL acquired are better. The quasi-opposition solution $QU_j = \{qu_{j,1}, qu_{j,2}, \dots, qu_{j,D}\}$ of U_j is the point between the opposite point and the center of search space.

$$qu_{j,k} = \begin{cases} \text{rand}(y_k, ou_{j,k}) & u_{j,k} \leq y_k \\ \text{rand}(ou_{j,k}, y_k) & u_{j,k} > y_k \end{cases} \tag{29}$$

where $y_k = \frac{u_{\min,k} + u_{\max,k}}{2}$, $\text{rand}(y_k, ou_{j,k})$ stands for a random number in $[y_k, ou_{j,k}]$, and $\text{rand}(ou_{j,k}, y_k)$ stands for a random number in $[ou_{j,k}, y_k]$. Finally, the initial population retains PN individuals having better solutions from $\{U_j \cup QU_j\}$.

4.3.2 The dynamic weight and nonlinear control parameter

The balance of the exploitation and exploration ability in WOA mostly lies on the coefficient A . The search range is extended for finding better solution if $|A| \geq 1$, which determines the exploration capability of WOA. The search range is narrowed for carrying out more careful search if $|A| < 1$, which determines the exploitation capability of WOA. By observing Eq. (10), it can be found that control parameter \vec{a} affects \vec{A} directly. However, \vec{a} decreases from 2 to 0 linearly during the course of iterations, which can not fully reflect the entire search process. Therefore, in this paper, nonlinear control parameter is proposed to improve convergence

performance of WOA, and it is formulated as

$$\vec{a} = 1 + \cos\left(\pi \times \frac{t}{MN}\right) \tag{30}$$

Figure 3 shows the changing trend of \vec{a} with the increase of the iteration number. It can be seen that nonlinear control parameter \vec{a} in the early stage is larger than original control parameter. It begins by the wide search range provided with high exploration ability. Moreover, nonlinear control parameter \vec{a} in the later stage is smaller than original control parameter, which makes the search range getting smaller provided with high exploitation ability. In this way, IWOA may keep an effective balance between the exploration and exploitation ability.

From Eq. (15), it can be seen that in the later local exploitation stage WOA can stay near the optima and not find the optima well. Inspired by particle swarm optimization (PSO) [46], the dynamic weight is proposed to perform a fine search near the optima, and it is formulated as

$$S = \left(1 - \frac{t}{MN}\right)^\mu \tag{31}$$

where $\mu > 0$, and μ is a constant coefficient used to adjust the decay extent of the dynamic weight. Correspondingly, mathematical expressions are updated below.

$$\vec{U}(t+1) = S \cdot \vec{U}_{\text{best}} - \vec{A} \cdot \vec{D} \tag{32}$$

$$\vec{U}(t+1) = \vec{D}' \cdot e^{bl} \cdot \cos(2\pi l) + S \cdot \vec{U}_{\text{best}} \tag{33}$$

From Eqs. (32) and (33), it can be seen that in the later stage the optima is more and more attractive to the whales. The dynamic weight is getting smaller and smaller so that the whales can find the optima more accurately, which can effectively improve optimization accuracy of WOA.

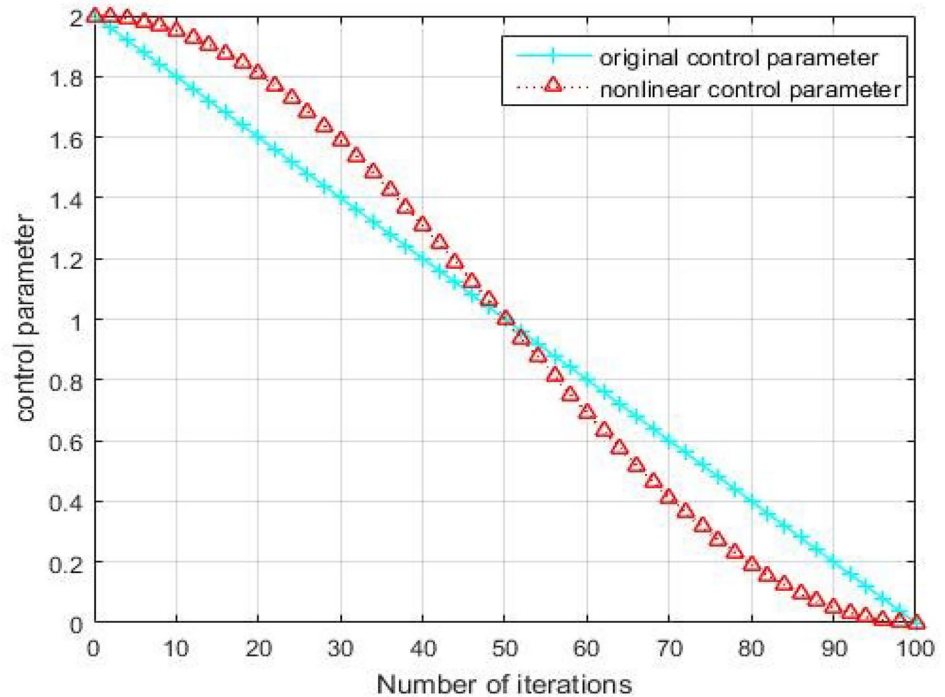
4.4 Selective ensemble

In this paper, selective ensemble based on binary version of IWOA is used for handling imbalanced dataset. All the base DWFELMs are expressed as binary strings by using a binary encoding form $(\alpha_1, \alpha_2, \dots, \alpha_i, \dots, \alpha_k)$, where k is the amount of base classifiers, $\alpha_i = 1$ means the i th base classifier is selected, and $\alpha_i = 0$ means the i th base classifier is discarded. Motivated by discrete PSO [47], binary IWOA is presented for selective ensemble of base classifier. In binary IWOA, the solution is transformed below.

$$\text{sigmoid}(u_{j,k}) = \frac{1}{1 + \exp(-u_{j,k})} \tag{34}$$

$$u'_{j,k} = \begin{cases} 1 & \text{if } \text{rand}(0, 1) < \text{sigmoid}(u_{j,k}) \\ 0 & \text{otherwise} \end{cases} \tag{35}$$

Fig. 3 The changing trend of control parameter



where $u'_{j,k}$ is the discretization solution. Then the integration by majority voting is adopted for the selected base DWFELMs.

Furthermore, the fitness of selective ensemble is formulated as

$$\text{fitness} = \frac{1}{s} \sum_{j=1}^s \mathfrak{N}(\hat{z}_j, z_j), \mathfrak{N}(\hat{z}_j, z_j) = \begin{cases} 1 & \text{if } \hat{z}_j = z_j \\ 0 & \text{if } \hat{z}_j \neq z_j \end{cases} \tag{36}$$

where \hat{z}_j is the target predicted concerning the j th testing sample, z_j is the target expected concerning the j th testing sample, and s is testing samples number. The ensemble performance is proportional to the fitness. The pseudo-code of IWOA is described in Fig. 4.

5 Experiments

5.1 Experimental datasets and experimental setting

To confirm the effectiveness of the proposed model, a number of comparative experiments are conducted on various gene expression data from GEMS repository [48]. The description of these datasets is listed in detail in Table 1.

The variation of these datasets attributes is quantitatively from 2000 to 12533, which are normalized into [0, 1]. The variation of these datasets classes is quantitatively from 2 to

11. The imbalance ratio (IR) is defined as

$$\begin{aligned} \text{Multi-class : } IR &= \frac{\max(\#z_j)}{\min(\#z_j)} \\ \text{Binary class : } IR &= \frac{\#majority}{\#minority} \end{aligned} \tag{37}$$

All the datasets are randomly divided into training–testing set. Based on a series of experiments, the results of classification performance are evaluated by the average of individually repeated 10 runs.

Experimental comparison of various approaches are made for evaluating the performance of SEN-DWFELM. All the experiments are test by using MATLAB platform and windows 10 OS with 8 GB of RAM. For an impartial comparison purpose, population size, the maximum iteration number and searching ranges for hidden nodes L , penalty parameter C are set the same for all the approaches. Set 40 for population size and 100 for the maximum iteration. Among comparative approaches, a grid search of hidden nodes L on $\{100, 110, \dots, 990, 1000\}$ and penalty parameter C on $\{2^{-18}, 2^{-16}, \dots, 2^{48}, 2^{50}\}$ is conducted, and $G(x) = \frac{1}{1+\exp(-(a \cdot x+b))}$ is used as activation function.

5.2 Measure metrics

Accuracy is the overall classification accuracy, which means the proportion of the correctly classified samples to all the samples. G-mean means the geometric mean of the pro-

Fig. 4 The pseudo-code of IWOA

```

Set the maximum number of iteration  $MN$  and the size of population  $PN$ 
The initial solutions are generated according to 4.3.1 and are transformed
by Eqs. (34) and (35)
Retain the optimal solution after calculating the fitness of every solution
while ( $t \leq MN$ )
  for  $j=1$  to  $PN$ 
    Update  $\vec{F}, \vec{A}, l$  and  $p$ 
    Update  $\vec{a}$  by Eq. (30)
    if ( $p < 0.5$ )
      if ( $|A| < 1$ )
        Modify the position of the present solution by Eq. (32)
      else if ( $|A| \geq 1$ )
        Choose a random solution
        Modify the position of the present solution by Eq. (8)
      end if
    else if ( $p \geq 0.5$ )
      Modify the position of the present solution by Eq. (33)
    end if
    Modify the solution if it does not within the search range
    The present solution is transformed by Eqs. (34) and (35)
  end for
  Calculate the fitness of every solution, then modify the optimal solution
   $t = t + 1$ 
end while
Return the optimal solution
    
```

Table 1 Description of these datasets

Datasets	#Atts	#Train	#Test	#Class	IR
Leukemia2	11225	38	34	3	1.4
Colon	2000	40	22	2	1.82
SRBCT	2308	43	40	4	2.64
DLBCL	5469	43	34	2	3.05
Leukemia1	5327	38	34	3	4.22
11_Tumors	12533	120	54	11	4.5

portion of the correct classification of all the classes. The higher G-mean is, the better classification performance of

every class is. F-measure is commonly used for evaluating class imbalanced problems, and F-score is the average of F-measure in all the classes. Based on above analysis, these measure metrics are expressed as follows.

$$\text{Accuracy} = \frac{\sum_{j=1}^c \text{TP}_j}{\sum_{j=1}^c \text{TP}_j + \text{FN}_j} \tag{38}$$

$$\text{SP}_j = \frac{\text{TP}_j}{\text{TP}_j + \text{FP}_j} \tag{39}$$

$$\text{SR}_j = \frac{\text{TP}_j}{\text{TP}_j + \text{FN}_j} \tag{40}$$

Table 2 Parameters setting

Datasets	ELM L	WELM1 (C,L)	WELM2 (C,L)	WELM (C,L)	DWFELM (C,L)	SEN-DWFELM (C,L)
Leukemia2	500	(2 ² ,590)	(2 ³² ,880)	(2 ³² ,910)	(2 ²⁸ ,800)	(2 ³⁰ ,200)
Colon	980	(2 ⁻² ,650)	(2 ⁰ ,990)	(2 ³⁶ ,750)	(2 ⁻² ,520)	(2 ²⁴ ,700)
SRBCT	600	(2 ¹⁴ ,550)	(2 ¹⁰ ,900)	(2 ⁻² ,930)	(2 ⁴⁰ ,890)	(2 ¹⁴ ,800)
DLBCL	650	(2 ⁰ ,920)	(2 ² ,940)	(2 ²² ,790)	(2 ²⁴ ,930)	(2 ³⁰ ,870)
Leukemia1	940	(2 ¹⁰ ,950)	(2 ¹⁸ ,870)	(2 ¹² ,560)	(2 ¹⁴ ,780)	(2 ¹⁰ ,700)
11_Tumors	940	(2 ² ,980)	(2 ² ,950)	(2 ⁻² ,960)	(2 ²⁶ ,890)	(2 ²⁶ ,790)

$$G\text{-mean} = \left(\prod_{j=1}^c SR_j \right)^{\frac{1}{c}} \quad (41)$$

$$F\text{-measure}_j = \frac{2SR_j \times SP_j}{SR_j + SP_j} \quad (42)$$

$$F\text{-score} = \frac{\sum_{j=1}^c F\text{-measure}_j}{c} \quad (43)$$

where TP_j is the number of the j th class samples correctly classified as the j th class samples, FN_j is the number of the j th class samples wrongly classified as other class samples, FP_j is the number of other class samples wrongly classified as the j th class samples, SP_j is precision ratio of the j th class samples, and SR_j is recall ratio of the j th class samples. Obviously, G-mean makes a more impartial comparison in that G-mean is 0 if classification accuracy of a certain class is 0 [49].

5.3 Comparison with variants of SEN-DWFELM

To evaluate classification performance of SEN-DWFELM, it is respectively compared with ELM, learning algorithm applying W1- weighted scheme (WELM1), learning algorithm applying W2-weighted scheme (WELM2) in this experiment. Meanwhile, SEN-DWFELM is also compared with its variants, namely learning algorithm applying W-weighted scheme (WELM) and DWFELM built to discuss the importance of different sections in SEN-DWFELM.

The detailed results of parameters setting for comparative algorithms are shown in Table 2, then the best parameters with regard to C and L are specified for corresponding datasets. The detailed results of G-mean, F-score, Accuracy and training time for comparative algorithms are shown in Table 3, then the best results are indicated in bold. From Table 3, it can be observed that SEN-DWFELM obtains better G-mean and F-score compared with other approaches. On Leukemia2, Colon, SRBCT, DLBCL, Leukemia1 and 11_Tumors datasets, compared with ELM, G-mean is improved by about 8.67%, 8.52%, 4.54%, 7.55%, 7.97% and 8.49%, then F-score is improved by about 8.45%, 7.63%, 4.32%, 7.28%, 7.20% and 8.10%. The reason is because ELM ignores minority class samples due to the

premise that the size of every class is relatively balanced. Compared with WELM1, WELM2 and WELM, G-mean obtained by SEN-DWFELM is improved by about 5.45%, 5.09% and 4.34% on average on all the datasets, then F-score obtained by SEN-DWFELM is improved by about 5.61%, 5.65% and 4.82% on average on all the datasets. The reason is because DWFELM is proposed by amending WELM in the study. In DWFELM, feature weighted fuzzy membership is presented to eliminate classification error from noise samples and improve generalization performance of ELM. Meanwhile, the weighted scheme is also presented to strengthen the impact of minority class samples on classification. Furthermore, DWFELM-based selective ensemble algorithm is proposed to improve classification performance and the stability of single DWFELM.

By observing Table 3, it can be found that SEN-DWFELM improves Accuracy remarkably on all the datasets. Accuracy is improved by about 6.70%, 5.25%, 5.24%, 4.28% and 3.31% on average compared with ELM, WELM1, WELM2, WELM and DWFELM. The standard deviation (SD) obtained by SEN-DWFELM is also much smaller than other comparative algorithms. It illustrates that SEN-DWFELM can maintain classification accuracy of majority class samples and improve classification accuracy of minority class samples. In short, generalization performance of SEN-DWFELM is superior to other competitors. Table 3 also shows training time obtained by measuring the average of 10 runs for different approach, which is used for evaluating the computation cost of comparative algorithms. From Table 3, it can be seen that SEN-DWFELM consumes more training time by learning multiple classifiers compared with other variants. It is acceptable because the proposed SEN-DWFELM is based on WELM with faster learning speed.

In order to analyze thoroughly the performance, F-measure of different approach for all the datasets is shown in Fig. 5, whose x -axis indicates the abbreviation for every class name to observe the results more distinctly. From the results, it can be observed that F-measure of BL class obtained by SEN-DWFELM on SRBCT dataset is worse than some other approaches. The reason is because F-measure of a certain class is on increase at the expense of a sharp decline in F-

Table 3 Performance results (Mean±SD) of different approaches

Datasets	Performance	ELM	WELM1	WELM2	WELM	DWFELM	SEN-DWFELM
Leukemia2	Accuracy(%)	89.12 ± 5.01	91.76 ± 4.76	91.18 ± 3.92	92.35 ± 3.16	92.94 ± 3.72	97.06 ± 2.77
	G-mean(%)	88.39 ± 5.25	90.77 ± 5.09	91.00 ± 4.28	92.28 ± 3.72	92.65 ± 4.20	97.06 ± 2.78
	F-score(%)	88.41 ± 5.15	90.74 ± 5.03	90.39 ± 4.14	91.41 ± 3.70	92.46 ± 3.79	96.86 ± 2.83
Colon	Training time(s)	0.0080 ± 0.00048	0.0050 ± 0.00033	0.0053 ± 0.00024	0.0049 ± 0.00026	0.0055 ± 0.00051	0.0065 ± 0.0013
	Accuracy(%)	85.00 ± 6.08	87.27 ± 5.59	87.27 ± 5.98	88.18 ± 5.34	89.09 ± 5.34	91.82 ± 3.59
	G-mean(%)	84.06 ± 6.97	87.37 ± 6.32	88.69 ± 5.90	87.70 ± 4.53	89.88 ± 4.51	92.58 ± 3.12
SRBCT	F-score(%)	83.39 ± 6.88	86.00 ± 6.00	86.28 ± 6.61	86.98 ± 5.26	88.31 ± 5.78	91.02 ± 4.48
	Training time(s)	0.0088 ± 0.00037	0.0041 ± 0.00044	0.0052 ± 0.00045	0.0038 ± 0.00035	0.0051 ± 0.00061	0.0132 ± 0.0045
	Accuracy(%)	93.25 ± 5.14	94.00 ± 4.74	94.25 ± 4.09	94.50 ± 3.50	96.00 ± 3.16	98.75 ± 1.32
DLBCL	G-mean(%)	94.27 ± 4.18	95.36 ± 3.77	95.10 ± 3.08	95.41 ± 2.98	96.74 ± 2.69	98.81 ± 1.33
	F-score(%)	94.18 ± 4.23	94.76 ± 4.22	94.91 ± 3.74	95.19 ± 3.19	96.56 ± 2.80	98.50 ± 1.76
	Training time(s)	0.0057 ± 0.00053	0.0015 ± 0.00013	0.0016 ± 0.00025	0.0012 ± 0.00013	0.0019 ± 0.00026	0.0070 ± 0.0021
Leukemia1	Accuracy(%)	92.94 ± 3.97	93.53 ± 3.87	93.24 ± 3.93	94.41 ± 3.52	95.59 ± 2.86	98.53 ± 1.55
	G-mean(%)	89.66 ± 6.25	92.42 ± 5.42	92.57 ± 6.16	94.23 ± 5.43	95.02 ± 3.82	97.21 ± 3.42
	F-score(%)	90.63 ± 5.08	91.82 ± 4.33	90.95 ± 4.95	91.80 ± 4.40	93.87 ± 4.34	97.91 ± 2.28
11_Tumors	Training time(s)	0.0075 ± 0.00024	0.0042 ± 0.00023	0.0047 ± 0.00089	0.0043 ± 0.00029	0.0048 ± 0.00036	0.0147 ± 0.0055
	Accuracy(%)	90.00 ± 5.41	92.06 ± 4.61	92.35 ± 4.21	93.53 ± 4.34	94.41 ± 3.78	97.65 ± 1.86
	G-mean(%)	90.19 ± 5.93	92.51 ± 4.88	92.74 ± 4.86	93.85 ± 4.57	94.57 ± 3.77	98.16 ± 1.56
11_Tumors	F-score(%)	90.91 ± 5.48	92.73 ± 4.78	93.16 ± 4.08	93.63 ± 4.06	94.71 ± 3.67	98.11 ± 1.51
	Training time(s)	0.0055 ± 0.00054	0.0026 ± 0.00021	0.0028 ± 0.00027	0.0023 ± 0.00019	0.0030 ± 0.00014	0.0098 ± 0.0033
	Accuracy(%)	89.07 ± 4.14	89.44 ± 4.54	89.81 ± 3.18	90.93 ± 2.95	91.67 ± 3.63	95.74 ± 1.96
11_Tumors	G-mean(%)	83.11 ± 5.87	84.27 ± 5.09	84.81 ± 4.26	85.94 ± 4.42	86.87 ± 4.21	91.60 ± 2.31
	F-score(%)	85.90 ± 5.06	86.72 ± 5.24	86.81 ± 4.45	88.45 ± 3.87	89.15 ± 2.85	94.00 ± 1.98
	Training time(s)	0.0210 ± 0.00031	0.0074 ± 0.00051	0.0076 ± 0.00083	0.0072 ± 0.00033	0.0108 ± 0.00054	0.0350 ± 0.0092

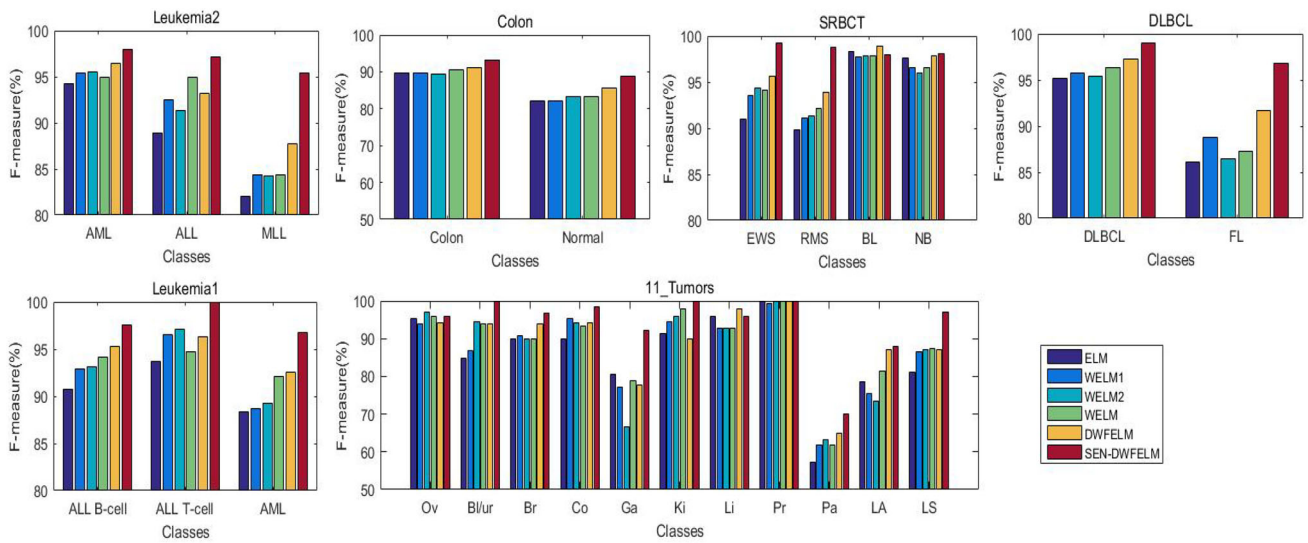


Fig. 5 F-measure of every class on all the datasets by using different approaches

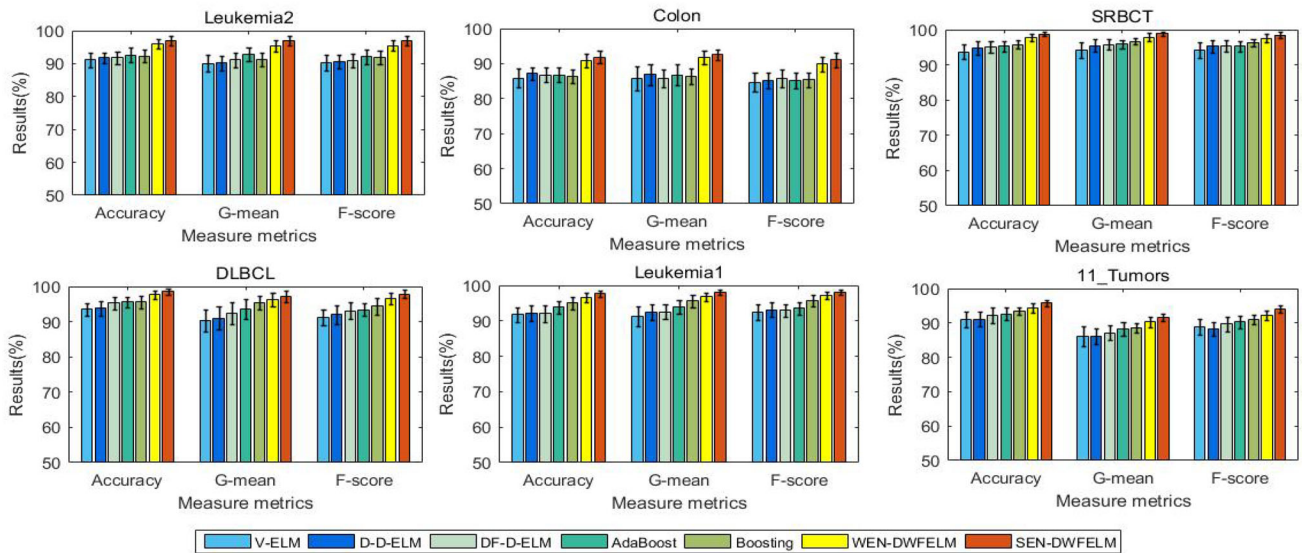


Fig. 6 Comparison of performance results

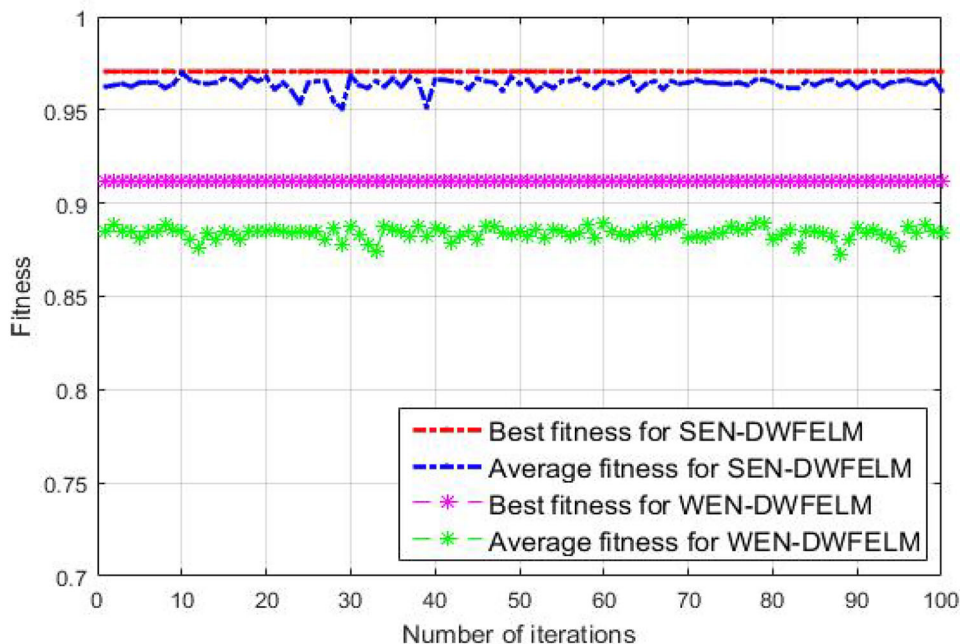
measures of other classes. From Fig. 5, it can be concluded that SEN-DWFELM can actually strengthen classification performance of minority class samples, and SEN-DWFELM is suitable for not only binary classification but also multi-class classification problems.

5.4 Comparison with other ensemble learning methods

SEN-DWFELM is also compared with other ensemble algorithms, namely V-ELM [13], DF-D-ELM, D-D-ELM [14], AdaBoost [11], Boosting [12] and WOA-based selective ensemble algorithm of DWFELM (WEN-DWFELM). Performance results in terms of G-mean, Accuracy, F-score and

SD are shown in Fig. 6. From Fig. 6, it can be observed that performance results acquired by SEN-DWFELM outperforms other comparative ensemble algorithms. Based on the dissimilarity measure, classification performance of DF-D-ELM and D-D-ELM is superior to V-ELM, and then classification performance of above approaches is also comparatively low. Especially on Colon and 11_Tumors dataset affected by complex data distribution, G-mean is significantly in decline. The reason is because D-D-ELM, V-ELM, and DF-D-ELM are all based on ELM algorithm in no consideration of weighted schemes. Therefore, these algorithms neglect minority class samples and then lead to the decline of G-mean. In AdaBoost, the importance of samples is indicated by the weight distributed for every training sample. When

Fig. 7 The best and average fitness obtained by WEN-DWFELM and SEN-DWFELM for run #1 on Leukemia1 dataset



samples are misclassified, the weights distributed for them are larger. On the contrary, the weights distributed for samples correctly classified are smaller. In Boosting, the weight distributed for every training sample from different class is updated apart in the light of classification performance of the former classifiers. Based on the weights distributed, classification performance of AdaBoost and Boosting is superior to D-D-ELM, V-ELM and DF-D-ELM. Compared to above comparative approaches, WEN-DWFELM performs better in terms of G-mean, Accuracy and F-score. It indicates the superiority of meta-heuristics algorithms based selective ensemble of classifiers. In all the circumstances, SEN-DWFELM obtains the best G-mean, Accuracy and F-score. The reason is because samples number of every class, feature weighted fuzzy membership and IWOA-based selective ensemble of classifiers are all taken into consideration for enhancing generalization performance of SEN-DWFELM. As shown in Fig. 6, dispersion degree of SEN-DWFELM in terms of G-mean, Accuracy and F-score is relatively lower, which illustrates the stability and robustness of the proposed model. From the analyses, it can be concluded that SEN-DWFELM is suitable for both imbalanced dataset and relatively balanced dataset.

To catch on the variation procedure of IWOA and WOA, Fig. 7 shows the evolution of the best and average fitness obtained by WEN-DWFELM and SEN-DWFELM for run #1 on Leukemia1 dataset. From this figure, it can be seen that the best fitness has no fluctuates during the evolution and the best fitness for SEN-DWFELM is larger than that for WEN-DWFELM. The average fitness has significant variation from iteration 1 to 100. The average fitness for SEN-DWFELM is

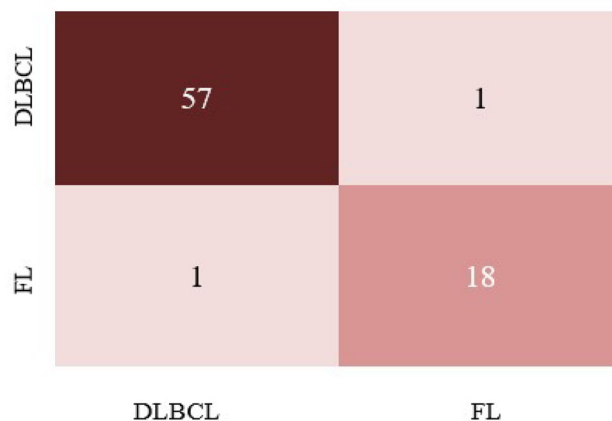


Fig. 8 The confusion matrix obtained by SEN-DWFELM on DLBCL dataset

close to the best fitness, and the average fitness for WEN-DWFELM is smaller than that for SEN-DWFELM. This phenomenon illustrates that the proposed IWOA can significantly improve the convergence rate and the quality of solution compared to WOA. The confusion matrix obtained by SEN-DWFELM on DLBCL dataset is shown in Fig. 8, where DLBCL denotes diffuse large B-cell lymphoma, and FL denotes follicular lymphoma. As observed from Fig. 8, SEN-DWFELM can classify 75 samples correctly in all 77 samples, where 1 DLBCL sample is misclassified as FL and 1 FL sample is misclassified as DLBCL. From the above results, it can be concluded that SEN-DWFELM can effectively improve diagnostic performance of tumor.

Table 4 Results of the paired t-test

Comparisons	<i>p</i> value
SEN-DWFELM versus ELM	0.0000160
SEN-DWFELM versus WELM1	0.0000055
SEN-DWFELM versus WELM2	0.0000048
SEN-DWFELM versus WELM	0.0000022
SEN-DWFELM versus DWFELM	0.0000530
SEN-DWFELM versus V-ELM	0.0000019
SEN-DWFELM versus D-D-ELM	0.0000049
SEN-DWFELM versus DF-D-ELM	0.0001300
SEN-DWFELM versus AdaBoost	0.0000710
SEN-DWFELM versus Boosting	0.0009600
SEN-DWFELM versus WEN-DWFELM	0.0000200

To further compare the performance, the paired t-test as the statistical testing method is utilized to study the difference among these comparative algorithms. Generally, the threshold is set as 0.05, and that the *p* value is less than 0.05 means the two methods are remarkably different. Then the *p* value results based on classification accuracy are shown in Table 4. As shown in Table 4, SEN-DWFELM exists evident difference with other comparative algorithms. Meanwhile, it can be seen that SEN-DWFELM is effective in coping with gene expression data.

6 Conclusions

Tumor classification is a complex task, which is closely related to the properties of multi-class imbalance, high noise and high-dimensional small samples of gene expression data. An effective SEN-DWFELM model is proposed for tumor classification using gene expression data in this paper. Feature weighted fuzzy membership is presented to eliminate classification error from noise samples, and then it reduces the dimensionality of sample by removing features with smaller weights to improve training efficiency. The weighted scheme is also designed to strengthen the relative impact of minority class samples and lessen the bias against performance from imbalanced dataset. Furthermore, meta-heuristic method based selective ensemble concept is developed for making classification performance more robust. The experiments are conducted on gene expression data of binary class and multi-class. Compared with its variants and conventional ensemble methods, experimental results prove that SEN-DWFELM significantly outperforms other competitors in terms of G-mean, Accuracy and F-score. In future, the proposed SEN-DWFELM model may help in practical medical diagnosis, and neural network in combination with machine

learning techniques can be applied to achieve better classification performance.

Acknowledgements This work was supported by talent scientific research fund of LIAONING PETROCHEMICAL UNIVERSITY (No.2023XJLL-006).

References

- Chen, W., Sun, K., Zeng, R., et al.: Cancer incidence and mortality in China, 2014. *Chin. J. Cancer Res.* **30**(1), 1–12 (2018)
- Kar, S., Sharma, K.D., Maitra, M.: Gene selection from microarray gene expression data for classification of cancer subgroups employing PSO and adaptive K-nearest neighborhood technique. *Expert Syst. Appl.* **42**(1), 612–627 (2015)
- Sun, L., Zhang, X.Y., Qian, Y.H., Xu, J.C., Zhang, S.G.: Feature selection using neighborhood entropy-based uncertainty measures for gene expression data classification. *Inf. Sci.* **502**, 18–41 (2019)
- Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: theory and applications. *Neurocomputing* **70**(1), 489–501 (2006)
- Wu, C., Li, Y.Q., Zhao, Z.B., Liu, B.: Extreme learning machine with autoencoding receptive fields for image classification. *Neural Comput. Appl.* **32**, 8157–8173 (2020)
- Wong, P.K., Huang, W., Vong, C.M., Yang, Z.X.: Adaptive neural tracking control for automotive engine idle speed regulation using extreme learning machine. *Neural Comput. Appl.* **32**, 14399–14409 (2020)
- Mohammed, A.A., Minhas, R., Wu, Q.M.J., Sid-Ahmed, M.A.: Human face recognition based on multidimensional PCA and extreme learning machine. *Pattern Recognit.* **44**(10), 2588–2597 (2012)
- Kaya, Y., Uyar, M.: A hybrid decision support system based on rough set and extreme learning machine for diagnosis of hepatitis disease. *Appl. Soft Comput.* **13**(8), 3429–3438 (2013)
- Lan, Y., Soh, Y.C., Huang, G.B.: Ensemble of online sequential extreme learning machine. *Neurocomputing* **72**(13), 3391–3395 (2009)
- Zhou, Z.H., Wu, J., Tang, W.: Ensembling neural networks: many could be better than all. *Artif. Intell.* **137**(1–2), 239–263 (2002)
- Shigei, N., Miyajima, H., Maeda, M., et al.: Bagging and AdaBoost algorithms for vector quantization. *Neurocomputing* **73**(1), 106–114 (2009)
- Li, K., Kong, X., Lu, Z., Liu, W., Yin, J.: Boosting weighted ELM for imbalanced learning. *Neurocomputing* **128**, 15–21 (2014)
- Cao, J.W., Lin, Z.P., Huang, G.B., Liu, N.: Voting based extreme learning machine. *Inf. Sci.* **185**(1), 66–77 (2012)
- Lu, H.J., An, C.L., Zheng, E.H., Lu, Y.: Dissimilarity based ensemble of extreme learning machine for gene expression data classification. *Neurocomputing* **128**, 22–30 (2014)
- Zhang, W.B., Ji, H.B.: Fuzzy extreme learning machine for classification. *Electron. Lett.* **49**(7), 448–449 (2013)
- He, H., Garcia, E.A.: Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **21**(9), 1263–1284 (2009)
- Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**(1), 321–357 (2002)
- Liu, X.Y., Wu, J., Zhou, Z.H.: Exploratory undersampling for class-imbalance learning. *IEEE Trans. Syst. Man Cybern. Part B* **39**(2), 539–550 (2009)
- Gupta, U., Gupta, D.: Bipolar fuzzy based least squares twin bounded support vector machine. *Fuzzy Set. Syst.* **449**, 120–161 (2022)

20. Hazarika, B.B., Gupta, D.: Density-weighted support vector machines for binary class imbalance learning. *Neural Comput. Appl.* **33**(9), 4243–4261 (2021)
21. Gupta, D.: Training primal K-nearest neighbor based weighted twin support vector regression via unconstrained convex minimization. *Appl. Intell.* **47**(3), 962–991 (2017)
22. Hazarika, B.B., Gupta, D.: Density weighted twin support vector machines for binary class imbalance learning. *Neural Process. Lett.* **54**(2), 1091–1130 (2022)
23. Hancer, E., Xue, B., Zhang, M.J.: Differential evolution for filter feature selection based on information theory and feature ranking. *Knowl.-based Syst.* **140**, 103–119 (2018)
24. Mirjalili, S., Lewis, A.: The whale optimization algorithm. *Adv. Eng. Softw.* **95**, 51–67 (2016)
25. Yan, Z.P., Zhang, J.Z., Zeng, J., Tang, J.L.: Nature-inspired approach: an enhanced whale optimization algorithm for global optimization. *Math. Comput. Simul.* **185**, 17–46 (2021)
26. Sun, Y.J., Wang, X.L., Chen, Y.H., Liu, Z.J.: A modified whale optimization algorithm for large-scale global optimization problems. *Expert Syst. Appl.* **114**, 563–577 (2018)
27. Fan, Q., Chen, Z.J., Li, Z., Xia, Z.H., Yu, J.Y., Wang, D.Z.: A new improved whale optimization algorithm with joint search mechanisms for high-dimensional global optimization problems. *Eng. Comput.* **37**(3), 1851–1878 (2021)
28. Wang, J.Z., Du, P., Niu, T., Yang, W.D.: A novel hybrid system based on a new proposed algorithm-multi-objective whale optimization algorithm for wind speed forecasting. *Appl. Energy* **208**, 344–360 (2017)
29. Aziz, M.A.E., Ewees, A.A., Hassanien, A.E.: Whale optimization algorithm and moth-flame optimization for multilevel thresholding image segmentation. *Expert Syst. Appl.* **83**, 242–256 (2017)
30. Mafarja, M.M., Mirjalili, S.: Hybrid whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing* **260**, 302–312 (2017)
31. Gao, L.Y., Ye, M.Q., Lu, X.J., Huang, D.B.: Hybrid method based on information gain and support vector machine for gene selection in cancer classification. *Genom. Proteom. Bioinf.* **15**(6), 389–395 (2017)
32. Rani, M.J., Devaraj, D.: Two-stage hybrid gene selection using mutual information and genetic algorithm for cancer data classification. *J. Med. Syst.* **43**(8), 235 (2019)
33. Tavasoli, N., Rezaee, K., Momenzadeh, M., Sehhati, M.: An ensemble soft weighted gene selection-based approach and cancer classification using modified metaheuristic learning. *J. Comput. Des. Eng.* **8**(4), 1172–1189 (2021)
34. Lu, H.J., Chen, J.Y., Yan, K., Jin, Q., Xue, Y., Gao, Z.G.: A hybrid feature selection algorithm for gene expression data classification. *Neurocomputing* **256**, 56–62 (2017)
35. Mondal, M., Semwal, R., Raj, U., Aier, I., Varadwaj, P.K.: An entropy-based classification of breast cancerous genes using microarray data. *Neural Comput. Appl.* **32**(7), 2397–2404 (2020)
36. Shukla, A.K., Singh, P., Vardhan, M.: Gene selection for cancer types classification using novel hybrid metaheuristics approach. *Swarm Evol. Comput.* **54**, 100661 (2020)
37. Dabba, A., Tari, A., Mefitali, S., Mokhtari, R.: Gene selection and classification of microarray data method based on mutual information and moth flame algorithm. *Expert Syst. Appl.* **166**, 114012 (2021)
38. Wang, Y., Wang, A.N., Ai, Q., Sun, H.J.: Enhanced kernel-based multilayer fuzzy weighted extreme learning machines. *IEEE Access* **8**, 166246–166260 (2020)
39. Wang, Y., Wang, A.N., Ai, Q., Sun, H.J.: An adaptive kernel-based weighted extreme learning machine approach for effective detection of Parkinson’s disease. *Biomed. Signal Process. Control* **38**, 400–410 (2017)
40. Bartlett, P.L.: The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. *IEEE Trans. Inf. Theory* **44**(2), 525–536 (1998)
41. Zong, W.W., Huang, G.B., Chen, Y.Q.: Weighted extreme learning machine for imbalance learning. *Neurocomputing* **101**(3), 229–242 (2013)
42. Palma-Mendoza, R.J., Rodriguez, D., De-Marcos, L.: Distributed ReliefF-based feature selection in spark. *Knowl. Inf. Syst.* **57**(1), 1–20 (2018)
43. Alotaibi, A.S.: Hybrid model based on ReliefF algorithm and k-nearest neighbor for erythematous-squamous diseases forecasting. *Arab. J. Sci. Eng.* **47**(2), 1299–1307 (2022)
44. Tizhoosh, R.H.: Opposition-based learning: A new scheme for machine intelligence. In: *International Conference on Computational Intelligence for Modelling, Control and Automation*, pp. 695–701 (2005)
45. Rahnamayan, S., Tizhoosh, H.R., Salama, M.M.A.: Quasi-oppositional differential evolution. In: *2007 IEEE Congress on Evolutionary Computation*, pp. 2229–2236 (2007)
46. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: *Proceedings of IEEE International Conference on Neural Network*, pp. 1942–1948 (1995)
47. Kennedy, J., Eberhart, R.C.: A discrete binary version of the particle swarm algorithm. In: *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, pp. 4104–4108 (1997)
48. <http://www.gems-system.org/>
49. Wang, Y., Wang, A.N., Ai, Q., Sun, H.J.: Ensemble based fuzzy weighted extreme learning machine for gene expression classification. *Appl. Intell.* **49**, 1161–1171 (2019)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.