

# 基于差异协作策略和稳定生长策略的青蒿素优化算法及其在高维医学特征选择问题的应用

姜永昆

温州大学计算机与人工智能学院, 浙江 温州

收稿日期: 2026年4月18日; 录用日期: 2026年5月14日; 发布日期: 2026年5月26日

## 摘要

高维医学数据普遍存在特征维度高、样本量小、冗余与噪声干扰等特征, 易造成计算负担过重、模型过拟合及诊断精度下降等问题, 特征选择是解决该问题的关键技术。元启发式算法在特征选择领域应用广泛, 但原始青蒿素优化算法在处理高维医学特征选择任务时, 易出现收敛停滞现象, 导致特征筛选效率偏低。为此, 文章提出一种融合差异协作策略与稳定生长策略的改进青蒿素优化算法, 通过个体间差异协作实现有效信息交互与协同寻优, 利用稳定生长策略维持种群多样性, 协同提升算法的全局搜索与局部优化能力, 进而增强其高维医学数据中的特征选择效率与筛选效果, 为医学数据挖掘与智能诊疗提供更高效特征优化方法。

## 关键词

元启发式算法, 特征选择, 青蒿素优化算法

## Artemisinin Optimization Based on Differential Collaboration and Stable Growth Strategies and Its Application in High-Dimensional Medical Feature Selection Problems

Yongkun Jiang

College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou Zhejiang

Received: April 18, 2026; accepted: May 14, 2026; published: May 26, 2026

文章引用: 姜永昆. 基于差异协作策略和稳定生长策略的青蒿素优化算法及其在高维医学特征选择问题的应用[J]. 计算机科学与应用, 2026, 16(5): 196-208. DOI: 10.12677/csa.2026.165176

## Abstract

High-dimensional medical data generally exhibits characteristics such as high feature dimensionality, small sample size, and interference from redundancy and noise, which easily lead to problems including excessive computational burden, model overfitting, and reduced diagnostic accuracy. Feature selection serves as a key technique to address these issues. Although metaheuristic algorithms have been widely applied in the field of feature selection, the original Artemisinin Optimization Algorithm is prone to convergence stagnation when handling high-dimensional medical feature selection tasks, resulting in low efficiency of feature screening. To this end, this paper proposes an improved Artemisinin Optimization Algorithm integrating a differential collaboration strategy and a stable growth strategy. It realizes effective information interaction and collaborative optimization through differential collaboration among individuals, maintains population diversity via the stable growth strategy, and synergistically enhances the global search and local optimization capabilities of the algorithm. Consequently, it improves the efficiency and performance of feature selection on high-dimensional medical data, providing a more efficient feature optimization method for medical data mining and intelligent diagnosis and treatment.

## Keywords

Metaheuristic Algorithm, Feature Selection, Artemisinin Optimization

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着基因芯片、蛋白质组学等技术的快速发展,医学数据呈现出样本量少、特征维度极高、冗余与噪声并存的典型特点。由于冗余与无关特征的存在,这类高维医学数据直接用于疾病诊断与分类模型时,不仅会大幅提升计算开销,还易引发过拟合与模型泛化能力下降等问题,严重制约临床辅助诊断的准确性与可靠性。直接处理此类数据不仅消耗大量算力,还会削弱模型诊断的准确性[1]。因此,针对高维医学数据的分析与处理而言,特征选择(Feature Selection, FS)显得至关重要。传统的医学FS方法,往往在面对高维小样本的基因表达数据时存在明显局限[2]。这类方法大多依赖人工设定规则或简单统计指标,难以捕捉特征间复杂的非线性关联,容易筛选出大量冗余甚至噪声特征,导致模型泛化能力下降。而元启发式优化算法凭借其灵活的搜索机制与较强的全局寻优能力,已成为解决高维FS问题的主流方法[3]。针对原始青蒿素优化算法(Artemisinin Optimization, AO)在高维医学特征选择中存在易出现收敛停滞,导致特征筛选效率过低的问题,本文提出了一种基于差异协作策略和稳定生长策略的改进青蒿素优化算法,通过个体间差异协作实现有效信息交互与协同寻优;同时借助稳定生长策略维持种群多样性,二者协同增强了算法的全局搜索与局部优化能力,进而提升算法在高维医学数据中的特征选择效率与效果。

## 2. 相关技术介绍

### 2.1. AO 算法简介

AO是受疟疾治疗过程启发的新型优化方法。疟原虫侵入人体红细胞并通过蚊子叮咬传播,模拟青蒿素治疗疟疾的三个关键阶段:全面消除阶段、局部清除阶段以及后巩固阶段[4]。

全面消除阶段：在疟疾治疗初期，患者需服用高剂量青蒿素以快速控制病情；药物被人体吸收后，会迅速在体内扩散至各组织部位。AO算法的全面消除阶段模拟这一过程，通过如下药剂扩散模型1实现全局搜索的快速遍历：

$$A_{i,j}^{t+1} = \begin{cases} A_{i,j}^t + c \times A_{i,j}^t \times (-1)^t, & r_1 < 0.5 \text{ and } r_2 < K \\ A_{i,j}^t + c \times best_j^t \times (-1)^t, & r_1 < 0.5 \text{ and } r_2 < K \end{cases} \quad (1)$$

其中， $t$ 为当前迭代次数， $r_1$ 、 $r_2 \in [0,1]$ ， $A_{i,j}$ 为第 $i$ 个搜索代理第 $j$ 维位置， $best_j$ 为最优搜索代理第 $j$ 维位置。随迭代推进，算法引入衰减因子 $c$ （模拟药物浓度降低），其计算见公式(2)：

$$c = 1 \times e^{-4 \times \frac{t}{Maxf}} \quad (2)$$

其中， $f$ 表示当前的评估次数，而 $Maxf$ 表示最大评估次数。考虑到不同患者对青蒿素的用药剂量需求及药物吸收率存在固有个体差异，公式(3)创新性引入概率因子 $K$ 。该因子作为算法的核心自适应参数，模拟临床治疗中针对个体差异的剂量调整策略，为AO算法的搜索过程注入了必要的适应性，以应对解空间中不同区域的搜索需求差异。

$$K = 1 - \frac{f^{1/6}}{Maxf^{1/6}} \quad (3)$$

局部消除阶段：尽管前述青蒿素治疗方案可有效缓解疟疾临床症状，但患者体内仍可能残留少量疟原虫。为彻底清除残余疟原虫，临床治疗中会采用持续低剂量青蒿素及其衍生物的局部清除策略[4]。AO算法模拟这一过程，通过公式(4)确定青蒿素药物粒子的新位置，实现解空间内的精细搜索：

$$A_{i,j}^{t+1} = A_{b_3}^t + d \times (A_{b_1}^t - A_{b_2}^t), \quad r_3 < Fit_{norm}(i) \quad (4)$$

$$b_1, b_2, b_3 \in \mathcal{U}(1, N), \quad b_1 \neq b_2 \neq b_3 \quad (5)$$

$$Fit_{norm}(i) = \frac{fit(i) - \min(fit)}{\max(fit) - \min(fit)} \quad (6)$$

其中， $d$ 为服从 $[0.1, 0.6]$ 区间均匀分布的随机数， $r_3$ 为 $[0,1]$ 区间内的均匀随机数；变量 $b_1$ 、 $b_2$ 、 $b_3$ 为 $[1, N]$ 范围内互不相同的整数，其选取规则如公式(5)所示。此外， $Fit_{norm}(i)$ 代表第 $i$ 个搜索代理的归一化适应度值，具体计算方式详见公式(6)。

后巩固阶段：在局部清除阶段结束后，尽管绝大多数疟原虫已被清除，但仍有少量疟原虫可能对青蒿素产生抗药性并进入休眠状态[4]。为模拟临床中防止疟疾复发的巩固治疗策略，式(7)提供了一种自适应搜索方法，用于检测解空间中这类休眠态最优解，确保算法能够收敛到全局最优解：

$$A_{i,j}^{t+1} = \begin{cases} A_{i,j}^t, & r_4 < 0.05 \\ best_j^t, & r_5 < 0.2 \end{cases} \quad (7)$$

其中， $r_4$ 、 $r_5$ 为服从 $[0,1]$ 区间均匀分布的随机变量； $best_j^t$ 表示第 $j$ 维的最优解。

## 2.2. 特征选择技术介绍

包裹式特征选择方法(Wrapper-Based FS) [5]是现阶段认可度更高的FS优化范式，其核心思路是把FS转化为组合优化寻优问题，直接将机器学习分类器的性能指标，也就是分类准确率作为特征子集的评价函数。依托在特征空间内的迭代搜索，该方法能够持续完成特征子集的筛选、评估与优化。相较于前两种FS方法，包裹式FS深度结合了分类器的反馈结果，可以更精准地筛选出对模型性能提升作用最大的特征组合，进而实现更优质的分类效果。

K-最近邻分类器(KNN) [6]是一类应用广泛的经典分类模型，其核心工作原理可概括为：针对预先给定的训练样本集合，对待判别样本进行分类判定时，首先筛选出与该样本空间距离最近的K个邻近样本，再依据多数投票原则完成分类；若选取的K个邻近样本里隶属于某一类别的样本占比最高，即可将待分类样本划归至该类别。在距离的度量方面，本文采用欧氏距离。两点之间的欧氏距离计算公式如式(8)所示：

$$\text{Distance} = \sqrt{\sum_{k=1}^n (M_k - N_k)^2} \quad (8)$$

其中，Distance代表了两个样本间的距离，M代表了训练样本，N代表测试样本，n为样本个数。

FS的核心思路是通过筛选规模最小的特征子集，进而达成最优的分类精度目标。整个流程一方面致力于优化分类结果，提升分类准确率，另一方面也能大幅压缩建模所需的特征维度，精简特征总量。从问题本质来看，FS属于典型的离散组合优化问题。基于此，本文采用二进制编码规则，以数字0代表剔除对应特征，数字1代表保留对应特征，完成特征状态的量化表示；在此基础上，对元启发式算法进行针对性改进与适配，保障算法可高效求解FS这类二进制优化场景下的问题。本文所使用的二进制转换函数如公式(9)和公式(10)所示：

$$\text{Tr}(x) = \frac{1}{e^{-10 \times (x-0.5)} + 1} \quad (9)$$

$$X_{i,j} = \begin{cases} 1 & \text{rand} < \text{Tr}(x) \\ 0 & \text{rand} > \text{Tr}(x) \end{cases} \quad (10)$$

其中， $X_{i,j}$ 为第*i*个离散个体在第*j*维度上的值； $\text{Tr}(x)$ 为修正的Sigmoid转换函数，将连续值映射为特征选择概率；rand为[0,1]区间内的随机数，用于判定特征的选取(1)或舍弃(0)。

本文将FS视为一个单目标问题来求解，具体的目标函数如公式(11)所示：

$$\text{Fitness} = \alpha \cdot \text{Error} + \beta \cdot \frac{|N|}{|M|} \quad (11)$$

其中，Error为分类器的分类准确率，|N|为所选特征子集的长度，|M|为数据集的特征总数。 $\alpha$ 和 $\beta$ 分别是分类错误率的权重以及所选特征子集的长度的权重，本研究将 $\alpha$ 设置为0.95， $\beta$ 设置为0.05。为降低实验的偶然误差，采用10折交叉验证评估模型性能。FS的完整执行流程如图1所示。

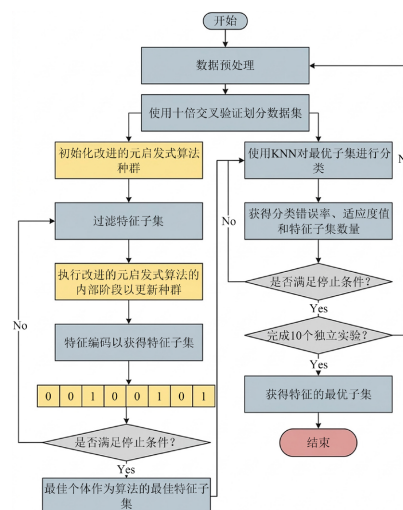


Figure 1. Flowchart of feature selection  
图1. 特征选择流程图

### 3. DOAO 算法

AO是一种模拟生物化学作用机理的新型启发式算法,其灵感源于自然界中青蒿素的提取过程或其抗疟治疗中的药理演化机制。然而,在处理复杂的优化问题及高维目标函数时,该算法存在全局勘探与局部开发失衡以及种群多样性不足的问题。为了弥补AO算法在这两方面的缺陷,本文引入了差异协作策略,通过增强搜索代理间的信息交互来拓宽种群的勘探范围,同时结合稳定生长策略,增强种群多样性以强化协同寻优过程中的自调节能力,优化全局探索与局部开发间的动态平衡。结合上述组合策略改进后的算法命名为基于差异协作策略和稳定生长策略的青蒿素优化算法(DOAO)。

#### 3.1. 差异协作策略

差异协作策略(Differential Collaboration Strategy, DS)的概念来源于差异化创新搜索(Differentiated Creative Search, DCS) [7],其核心是基于差分进化(DE)算法的current-to-pbest框架[8],通过重构变异操作实现了更精细的群体协作与搜索调控。其核心思想是让当前个体在吸收团队知识的过程中,既能借鉴最优个体的核心经验,又能灵活整合随机队友的差异化信息,从而在探索与收敛之间取得平衡。DE算法的current-to-pbest由公式(12)表示:

$$V_i = F \times X_{best} - (F \times X_{r_2} - X_i) + F \times (X_{r_1} - X_i) \quad (12)$$

其中,  $V_i$  表示第*i*个个体的变异向量,  $F$  表示缩放因子,  $X_{best}$  表示当前种群中的最优个体,  $X_{r_1}$ 、 $X_{r_2}$  表示从种群中随机选取的,与  $X_i$  互不相同的个体,  $X_i$  表示第*i*个目标向量。

通过引入三个独立调控系数,实现了对协作行为的精准控制:个体  $X_i$  整合团队成员  $X_{r_1}$  和  $X_{r_2}$  提供的信息,对团队核心(最优个体)  $X_{best}$  的知识进行优化完善。基于此,公式(12)可重新推导为如下通用形式:

$$V_i = w \times X_{best} + \lambda_t \times (X_{r_2} - X_i) + \omega_{i,t} \times (X_{r_1} - X_i) \quad (13)$$

这一全新的通用形式通过引入三个独立调控系数,实现了对协作行为的精准控制:  $w$ 、 $\lambda_t$  和  $\omega_{i,t}$ , 每个系数在调节当前向量与随机向量的差值贡献方面承担不同作用。权重因子  $w$  用于调控最优向量( $X_{best}$ )的影响程度。其默认值为1,但不同取值会改变候选个体在借鉴团队核心(最优个体)知识时的行为模式:若  $w \neq 1$ , 个体可能偏离核心的策略框架,转而采用创新性方法。 $\lambda_t$  代表第*t*次迭代时个体的  $\lambda$  系数值,该系数的计算公式如下:

$$\lambda_t = 0.1 + 0.518 \times \left(1 - \sqrt{FEs_t / FEs_{max}}\right) \quad (14)$$

其中,  $FEs_t$  表示第*t*次函数迭代次数,  $FEs_{max}$  表示算法允许的最大函数迭代次数。

差异协作策略通过动态平衡“核心经验借鉴”与“差异化信息整合”,让群体在协作优化过程中既保持高效收敛能力,又具备足够的多样性以避免陷入局部最优,增强了原始AO算法的性能。

#### 3.2. 稳定生长策略

稳定生长策略(Ordered Population Growth, OG)受自然界常春藤攀附生长行为启发[9],将种群个体模拟为藤蔓,通过初始化生长、局部攀附、全局收敛的递进流程实现优化求解。

设种群个体数量为  $N_{pop}$ , 问题决策变量维度为  $D$ , 则第*i*个个体可表示为  $I_i = (I_{i1}, \dots, I_{iD})$ , 整个种群记为  $\bar{I} = (I_1, \dots, I_{N_{pop}})$ 。在初始化阶段,种群个体在预设搜索空间内按公式(15)随机生成。

$$I_i = I_{min} + \text{rand}(1, D) \odot (I_{max} - I_{min}), \quad i = 1, \dots, N_{pop} \quad (15)$$

其中,  $\text{rand}(1, D)$  为  $[0, 1]$  均匀分布的  $D$  维随机向量,  $I_{max}$ 、 $I_{min}$  为搜索空间上下界,  $\odot$  表示向量按元素乘

积。

为刻画藤蔓的动态生长特性，将生长速率  $G_v(t)$  建模为时间的函数，其变化规律由公式(16)描述。

$$\frac{dG_v(t)}{dt} = \psi \cdot G_v(t) \cdot \varphi(G_v(t)) \quad (16)$$

其中， $\psi$  为控制系数， $\varphi(G_v(t))$  为生长偏差修正因子。

结合迭代过程，生长速率变化量可离散为差分形式，如公式(17)所示：

$$\Delta Gv_i(t+1) = \text{rand}^2 \odot (N(1, D) \odot \Delta Gv_i(t)) \quad (17)$$

其中， $\text{rand}^2$  为服从特定概率密度的随机数， $N(1, D)$  为标准高斯分布向量。

在局部攀附阶段，对种群按适应度排序得到  $\bar{I}^s = [I_1^s, \dots, I_{N_{pop}}^s]$ ，最优个体为  $I_{Best} = I_1^s$ ，邻居定义为公式(18)所示：

$$I_{ii} = \begin{cases} I_{j-1}^s, & I_i = I_j^s; \\ I_i, & I_i = I_{Best}. \end{cases} \quad (18)$$

个体沿邻居方向向更优解移动，位置更新规则为公式(19)：

$$I_i^{new} = I_i + |N(1, D)| \odot (I_{ii} - I_i) + N(1, D) \odot \Delta Gv_i \quad (19)$$

其中，生长速率变化量随迭代阶段动态调整，如公式(20)所示：

$$\Delta Gv_i = \begin{cases} I_i \odot (I_{\max} - I_{\min}), & \text{Iter} = 1; \\ \text{rand}^2 \odot (N(1, D) \odot \Delta Gv_i), & \text{Iter} > 1, \end{cases} \quad (20)$$

其中， $\odot$  为按元素除法， $|N(1, D)|$  表示对高斯向量各分量取绝对值。

局部搜索完成后，算法进入全局收敛阶段，个体围绕最优个体  $I_{Best}$  搜索如公式(21)所示：

$$I_i^{new} = I_{Best} \odot (\text{rand}(1, D) + N(1, D) \odot \Delta Gv_i) \quad (21)$$

新生个体的生长速率同步更新为如公式(22)所示：

$$\Delta Gv_i^{new} = I_i^{new} \odot (I_{\max} - I_{\min}) \quad (22)$$

稳定生长策略通过局部邻域探索(局部攀附)与全局最优开发(全局收敛)相结合的两阶段机制，在保持种群多样性的同时实现高效收敛，并借助动态生长速率平衡了算法的探索与开发能力。

### 3.3. DOAO 算法结构

DOAO算法在原始AO基础上，融合DS与OG两种策略。在每一轮迭代过程中，算法会先执行AO原始算法中全面清除和局部清除两个阶段，然后执行DS，借助重构变异的方式，对群体协作进行更加精细的调节与控制。该策略引入三个相互独立的调节参数，以此构建通用的变异方式，让个体在更新过程中，既能借鉴当前最优个体的搜索经验，也能结合其他随机个体与自身的差异信息进行调整。其中  $\lambda_i$  会随着函数迭代次数的增加逐步动态减小，以此精准平衡对最优个体经验的借鉴和对其他个体差异化信息的整合，使个体在吸收优秀经验的同时，还能灵活融合不同随机个体的差异化信息，进而提升整个种群的搜索多样性与收敛过程的稳定性，有效改善原有算法容易陷入局部最优解的缺陷。随后执行OG：将种群个体模拟为常春藤藤蔓，局部攀附阶段对种群按适应度排序，个体沿邻居方向向更优解移动，借助动态生长速率  $G_v(t)$  实现邻域精细探索，生长速率变化量  $\Delta Gv_i$  随迭代阶段动态调整，全局收敛阶段，个体围绕最优个体  $I_{Best}$  开展定向搜索，同步更新新生个体的生长速率，通过两阶段机制在保持种群多样性的同时实

现高效收敛，借助动态生长速率平衡算法的探索与开发能力。每轮迭代结束后，算法更新种群并重新评估所有个体适应度，同步更新全局最优个体  $I_{Best}$  与函数迭代次数  $FES$ ，直至达到最大函数迭代次数，最终输出全局最优解及其适应度值。该结构既通过DS实现了精准的群体协作与信息整合，又借助OG构建了稳定的动态生长与收敛机制，有效弥补了原始AO算法在动态探索开发平衡及收敛稳定性上的不足。

## 4. 基于差异协作策略和稳定生长策略的青蒿素优化算法对比实验

### 4.1. 实验设置

本次实验所用硬件环境搭载Intel (R) Core (TM) i5-14600KF中央处理器(主频3.50 GHz)，搭配32 GB运行内存，操作系统采用Windows 11 24H2版本，全部实验代码均依托MATLAB R2021a软件完成编写与运行。为全面验证DOAO算法的综合性能，选用IEEE 2017年进化计算大会(2017 Congress on Evolutionary Computation, CEC 2017)基准测试函数集开展性能测试。为保障对比实验的公平客观性，所有算法均在统一实验参数下运行：各算法的种群规模均设置为30，单次实验独立运行30次，测试问题维度设定为30维，算法最大适应度评估次数设为30万次，基准测试函数相关参数均沿用官方默认配置，各算法的相关参数均采用默认配置。将所提算法与十种元启发式算法进行横向对比实验。

### 4.2. DOAO 对比实验

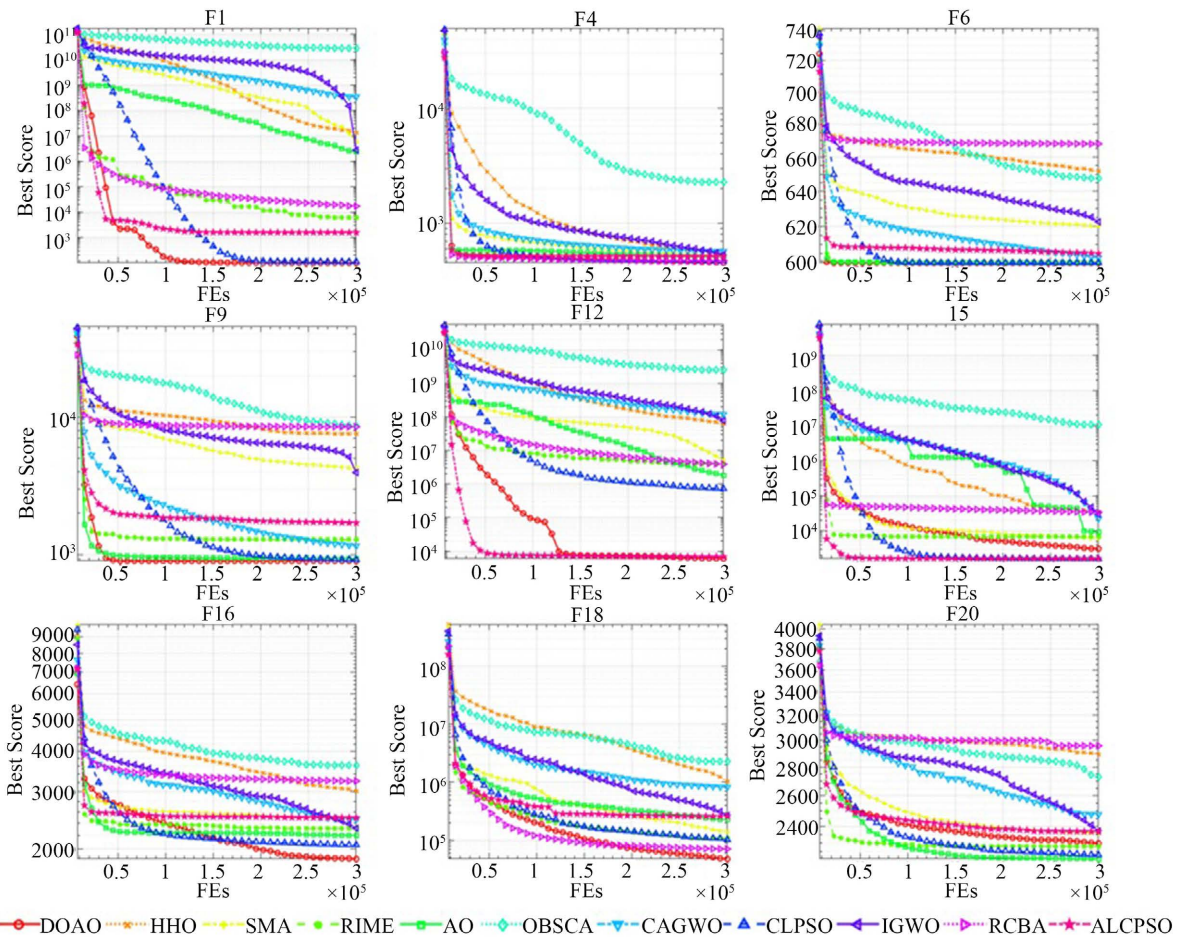


Figure 2. Convergence plots of DOAO and ten meta-heuristic algorithms on CEC 2017 benchmark functions

图2. DOAO与十种元启发式算法在CEC 2017基准函数上的收敛图

对比算法选择了十种元启发式算法：HHO [10]、SMA [11]、RIME [12]、AO [4]、OBSCA [13]、CAGWO [14]、CLPSO [15]、IGWO [16]、RCBA [17]、ALCPSO [18]，所有涉及的算法都在相同条件下进行了比较，实验中的公共参数进行统一设置。实验结果图2所示。从图中可以观察到，DOAO平均求解结果最小，且DOAO的结果非常接近对应的最优结果。这表明DOAO的综合优化性能良好，在单峰、多峰以及混合函数上性能比较突出，但在某些复合函数上效果一般。

## 5. 基于差异协作策略和稳定生长策略的青蒿素优化特征选择方法在高维医学数据集上的实验

本文将经过二进制转换的DOAO命名为bDOAO，再将bDOAO与KNN分类器结合，形成基于复杂变异策略和梯度搜索策略的龙格库塔优化特征选择方法，即bDOAO-KNN特征选择方法。

### 5.1. 实验设置

为全面评估bDOAO在高维FS问题中的性能，将其与bAO、bPSO [19]、bBA [20]、bWOA [21]以及bSCA [22]等二进制优化算法进行对比。为保持实验的一致性和公正性，所有比较算法均在相同条件下进行：种群规模设为20，问题的维度由数据集的特征数量确定，迭代次数固定为50。在进行比较时，所有算法的核心参数均保持一致，唯一的变化是这些算法均被适配为处理FS任务的二进制版本。采用10折交叉验证方法对UCI高维数据集和SBCB高维微阵列数据集展开实验进行分割，并以这10次交叉验证的平均结果作为评估算法FS性能的标准。此外，利用平均适应度、所选择特征的平均数量、分类的平均错误率和方法的平均运行时间四个指标来评价分类性能。这四项评价指标计算如公式(23)~(26)所示。

$$\text{Fitness} = \frac{1}{K} \sum_{i=1}^K \text{fitness}(i) \quad (23)$$

$$\text{Error} = \frac{1}{K} \sum_{i=1}^K (1 - \text{Accuracy}(i)) \quad (24)$$

$$\text{Features\_num} = \frac{1}{K} \sum_{i=1}^K \text{Size}(i) \quad (25)$$

$$\text{Time} = \frac{1}{K} \sum_{i=1}^K \text{Runtime}(i) \quad (26)$$

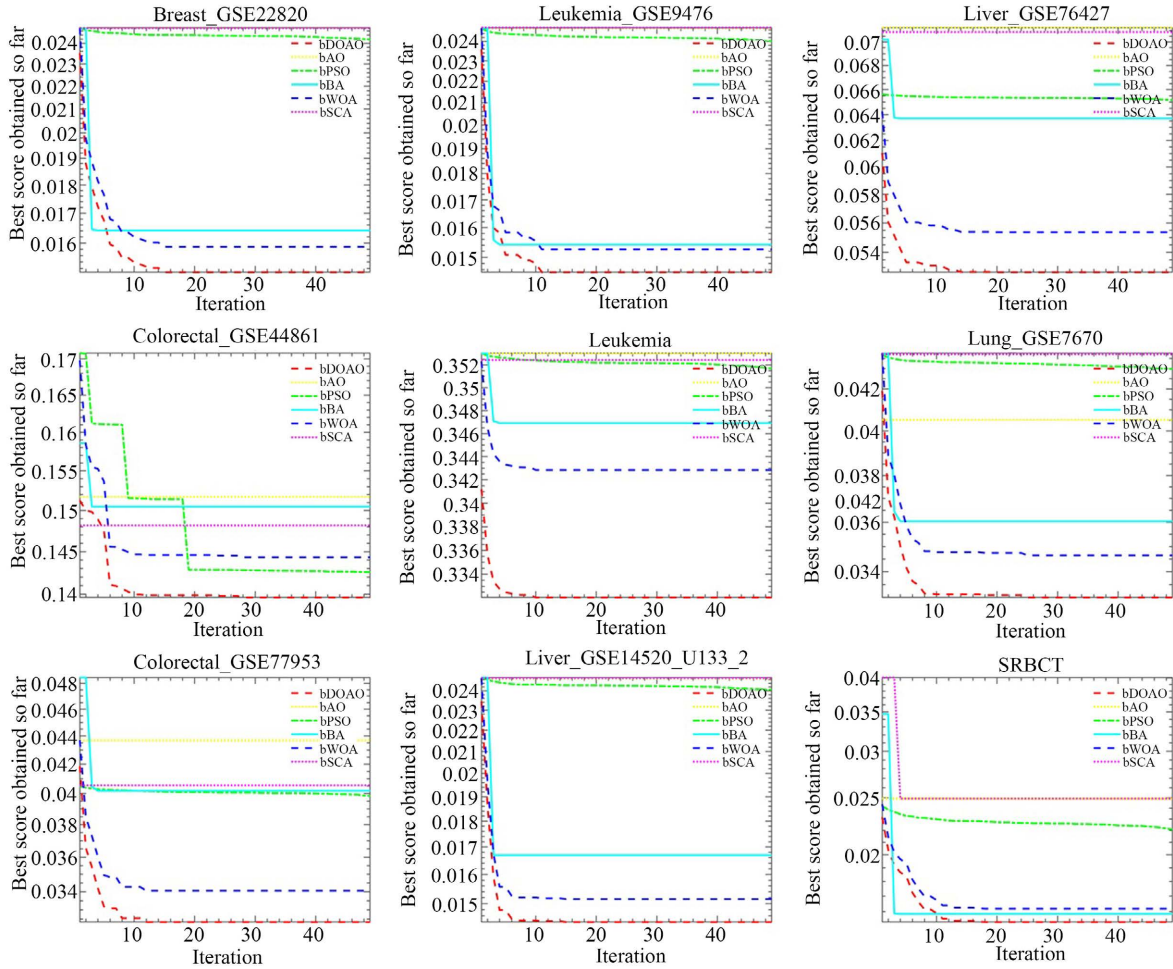
其中，

- $K$ ：交叉验证的折数(如10折交叉验证时  $K = 10$ )；
- $\text{fitness}(i)$ ：第*i*折验证的适应度值；
- $\text{Accuracy}(i)$ ：第*i*折验证的分类准确率；
- $\text{Size}(i)$ ：第*i*折验证中所选特征子集的大小；
- $\text{Runtime}(i)$ ：第*i*折验证的算法运行时间(单位：s)。

### 5.2. 数据集描述

实验采用UCI数据集以及SBCB的CuMiDa数据库中微阵列数据集开展。从UCI数据集中选取的2个高维数据集，包含Leukemia、SRBCT这两种医学基因表达数据；从SBCB微阵列数据集中选取7个高维微阵列数据集进行测试，包含Breast\_GSE22820、Colorectal\_GSE44861、Colorectal\_GSE77953、Leukemia\_GSE9476、Liver\_GSE14520\_U133\_2、Lung\_GSE7670、Liver\_GSE76427。

### 5.3. 特征选择实验结果与分析



**Figure 3.** Convergence plots of bDOAO and other algorithms on nine high-dimensional medical gene expression datasets  
**图3.** bDOAO与其他算法在九个高维医学基因表达数据集上的收敛图

各算法的适应度平均值详见表1，所选特征子集的平均数量列于表2，平均错误率呈现在表3中，算法的平均运行时间则见表4。从表中数据和各算法的收敛图(见图3)可以清晰地看出，在全部9个数据集上，bDOAO的适应度值表现均优于其他所有算法，这得益于差异协作策略和稳定生长策略的改进。在所选特征自己的平均大小方面，首先，由表2可知，在Leukemia和SRBCT这两个UCI数据集上，bDOAO可以将特征子集的平均大小降低至17.1和13.6，而其余方法则只能保留到100左右，在其余SBCB微阵列数据集中bDOAO也有很大的优势，这表明bDOAO在减少特征数量的同时，保持了分类的准确性。再看平均错误率表3，在Leukemia和SRBCT等低样本二分类问题中，bDOAO实现了出色的0错误率，在Leukemia\_GSE9476和Colorectal\_GSE77953等多分类问题中，bDOAO分别将数据集特征数量减少至约80个，而其他算法仅能缩小至一千多个，仅次优的bWOA做到了减少至一千个以内，这说明基于包裹式的FS方法能够通过全面考虑特征之间的相互关系并优化最终预测模型，选择出可使模型性能最大化的特征子集，大幅减少了数据集的特征数量。特别的是，与前两章提出的算法不同，在平均运行时间的表4中，bDOAO的时间开销做到了最小，说明其在保证特征选择精度的同时，展现出更优的计算效率，不仅能有效降低错误率和选择最优特征组合，还能在保证解决方案质量的同时控制运行时间。单独来看，将bDOAO与原始bAO相比，改

进后的bDOAO在全部四个指标方面均有提升,尤其是在全部数据集上的平均适应度值和所选特征子集的大小均明显小于bAO,在平均错误率上除在SRBCT和Colorectal\_GSE44861数据集上二者都为0外,其余七个数据集均小于bAO,在两种优化策略的改进下,时间开销方面不仅没有增加,反而比原始bAO有所减少。

综合来看,bDOAO算法在核心性能与时间开销方面均实现了领先,在高维医学基因表达数据的特征选择场景中表现出良好的综合竞争力,可为生物信息学领域的特征选择与疾病诊断研究提供更高效可靠的技术方案。

**Table 1.** Average fitness values of bDOAO and other algorithms datasets

**表1.** bDOAO与其他算法的平均适应度值

Datasets	Metric	bDOAO	bAO	bPSO	bBA	bWOA	bSCA
Breast_GSE228	STD	1.2702E-02	2.1225E-01	3.3612E-01	1.6088E-01	3.2016E-01	1.9749E-01
20	AVG	<b>1.5531E-02</b>	2.4189E-02	2.4371E-02	1.6319E-02	1.6193E-01	2.4108E-02
Colorectal_GS	STD	2.5771E-02	2.3687E-01	2.3364E-01	1.0772E-01	2.3140E-01	1.6010E-01
E44861	AVG	<b>1.4064E-01</b>	1.5308E-01	1.4505E-01	1.5302E-01	1.4686E-01	1.4821E-01
Colorectal_GS	STD	1.6351E-02	1.9046E-01	3.1782E-01	1.9465E-01	2.0622E-01	2.3632E-01
E77953	AVG	<b>3.3200E-02</b>	4.4298E-02	4.2046E-02	4.3706E-02	3.4521E-02	4.1075E-02
Leukemia	STD	1.9779E-02	2.8918E-01	2.0279E-01	1.5189E-01	1.8025E-01	1.8546E-01
	AVG	<b>3.3307E-01</b>	3.5314E-01	3.5228E-01	3.4780E-01	3.4303E-01	3.5160E-01
Leukemia_GSE	STD	1.0057E-02	1.7790E-01	2.1742E-01	2.0529E-01	3.7698E-01	1.2725E-01
9476	AVG	<b>1.4871E-02</b>	2.5297E-02	2.4333E-02	1.5854E-02	1.5702E-02	2.5877E-02
Liver_GSE1452	STD	1.6862E-02	1.9811E-01	2.4459E-02	1.8919E-02	2.8811E-02	9.4632E-02
0_U133_2	AVG	<b>1.4768E-02</b>	2.4864E-02	2.3996E-02	1.6417E-02	1.5812E-02	2.5982E-02
Lung_GSE7670	STD	4.3806E-02	2.2996E-01	4.9056E-01	9.5589E-02	2.2399E-01	2.2466E-01
	AVG	<b>3.3412E-02</b>	4.0543E-02	4.3171E-02	3.6182E-02	3.5075E-02	4.4134E-02
SRBCT	STD	8.2649E-02	1.6315E-01	2.0167E-01	2.0859E-01	2.8431E-01	1.7347E-01
	AVG	<b>1.0661E-02</b>	2.5079E-02	2.4627E-02	1.1261E-02	1.2566E-02	2.5933E-02
Liver_GSE7642	STD	5.9197E-02	1.1334E-01	2.7213E-01	1.9823E-01	3.1295E-01	1.3521E-01
7	AVG	<b>5.3523E-02</b>	7.1720E-02	6.6486E-02	6.4790E-02	5.6160E-02	7.1910E-02
	AVG	<b>1.00</b>	5.11	4.11	2.89	2.44	4.33
	Rank	<b>1</b>	6	4	3	2	5

**Table 2.** Average size of the feature subsets selected by bDOAO and other algorithms

**表2.** bDOAO和其他算法所选特征子集的平均大小

Datasets	Metric	bDOAO	bAO	bPSO	bBA	bWOA	bSCA
Breast_GSE228	STD	180.4284	1155.7343	3015.4624	217.4665	744.5710	1187.4131
20	AVG	<b>176.8</b>	1290.7	1153.7	952.4	656.1	1036.1
Colorectal_GS	STD	10.5798	234.9949	201.8246	1392.0715	629.1884	228.8324
E44861	AVG	<b>82.8</b>	814.2	1585.4	2460.8	818.0	965.8
Colorectal_GS	STD	219.8803	1389.5413	106.9187	920.5764	24.0391	964.1616
E77953	AVG	<b>79.1</b>	1384.6	1291.5	1318.1	424.1	1425.6
Leukemia	STD	17.8725	129.5167	107.5479	25.0514	94.4807	92.4859

续表

	AVG	<b>17.1</b>	104.7	119.5	98.4	61.7	118.0
Leukemia_GSE	STD	116.9785	19.8108	1392.0440	275.6847	391.8748	115.1365
9476	AVG	<b>84.2</b>	693.8	1164.4	874.1	696.5	1191.5
Liver_GSE1452	STD	30.0570	697.1237	1368.9096	706.0919	282.3280	1271.4152
0_U133_2	AVG	<b>91.1</b>	625.3	1154.1	701.5	881.5	1191.0
Lung_GSE7670	STD	119.0889	41.1092	1459.4344	374.4762	79.2140	1169.5414
	AVG	<b>70.6</b>	689.1	1161.6	769.6	723.3	1029.8
SRBCT	STD	11.9683	127.7269	144.7027	83.1348	92.2445	93.6460
	AVG	<b>13.6</b>	58.5	166.3	109.9	89.7	109.4
Liver_GSE7642	STD	49.8819	119.2675	423.5947	105.9437	500.1140	110.0104
7	AVG	<b>108.2</b>	700.6	1154.4	846.0	382.3	1099.4
	AVG	<b>1.00</b>	3.78	5.56	4.44	2.44	4.67
	Rank	<b>1</b>	3	6	4	2	5

**Table 3.** Average error rates of bDOAO and other algorithms**表3.** bDOAO与其他算法的平均错误率

Datasets	Metric	bDOAO	bAO	bPSO	bBA	bWOA	bSCA
Breast_GSE228	STD	1.9994E-01	7.9058E-02	1.4639E-01	1.7607E-01	4.0350E-01	1.6092E-01
20	AVG	<b>1.8000E-05</b>	2.5005E-02	1.0895E-01	1.3500E-01	3.8749E-02	1.2250E-02
Colorectal_GS	STD	2.3676E-01	1.0285E-01	2.1573E-01	2.8923E-01	2.4966E-01	1.9797E-01
E44861	AVG	<b>4.4715E-04</b>	2.7495E-01	4.1485E-01	4.1355E-01	4.0857E-02	3.5855E-01
Colorectal_GS	STD	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00
E77953	AVG	<b>0.0000E+00</b>	<b>0.0000E+00</b>	<b>0.0000E+00</b>	<b>0.0000E+00</b>	<b>0.0000E+00</b>	<b>0.0000E+00</b>
Leukemia	STD	0.0000E+00	7.9075E-02	2.176E-01	1.7495E-01	3.4255E-01	1.7215E-01
	AVG	<b>0.0000E+00</b>	2.5005E-02	2.100E-01	1.0005E-02	4.8786E-02	9.2857E-02
Leukemia_GSE	STD	1.6131E-02	1.2495E-01	1.4415E-01	1.4355E-01	2.0536E-01	1.5264E-01
9476	AVG	2.9286E-02	9.6429E-02	<b>2.6785E-02</b>	2.7125E-01	3.8571E-02	1.6428E-01
Liver_GSE1452	STD	2.0403E-01	1.6239E-01	2.236E-01	1.1735E-01	0.0000E+00	1.5034E-01
0_U133_2	AVG	3.4728E-04	2.1273E-01	3.0775E-02	3.2796E-02	<b>0.0000E+00</b>	2.5905E-02
Lung_GSE7670	STD	2.2117E-01	1.0545E-01	2.2065E-01	2.9499E-01	2.1726E-01	2.2965E-01
	AVG	<b>2.3505E-03</b>	4.9995E-02	1.9500E-01	1.6605E-02	3.2679E-02	1.9500E-01
SRBCT	STD	0.0000E+00	0.0000E+00	5.9292E-02	5.9292E-02	1.9365E-02	6.7765E-02
	AVG	<b>0.0000E+00</b>	<b>0.0000E+00</b>	1.875E-02	1.875E-02	1.5000E-02	2.1429E-02
Liver_GSE7642	STD	1.16565E-03	9.0351E-02	1.5655E-01	1.4985E-01	1.6065E-01	1.5495E-01
7	AVG	<b>8.57145E-06</b>	4.2865E-02	7.2865E-02	8.5745E-02	1.1785E-01	4.9995E-02
	AVG	<b>1.11</b>	3.33	4.33	3.44	2.67	3.22
	Rank	<b>1</b>	4	6	5	2	3

**Table 4.** Average runtime of bDOAO and other algorithms  
**表4.** bDOAO和其他算法的平均运行时间

Datasets	Metric	bDOAO	bAO	bPSO	bBA	bWOA	bSCA
Breast_GSE228	STD	3.5472E+01	6.8295E+01	4.2634E+01	5.7827E+01	7.3351E+01	2.6698E+01
20	AVG	1.5264E+02	1.9457E+02	1.5784E+02	1.8250E+02	<b>1.3527E+02</b>	1.6972E+02
Colorectal_GS	STD	4.2638E+01	7.9572E+01	4.5843E+01	6.2496E+01	3.1269E+01	5.8724E+01
E44861	AVG	<b>1.1856E+02</b>	1.8927E+02	1.6250E+02	1.7464E+02	1.9681E+02	1.2975E+02
Colorectal_GS	STD	2.7295E+01	5.5618E+01	7.8347E+01	4.3762E+01	3.6589E+01	6.8426E+01
E77953	AVG	<b>1.3174E+02</b>	1.9366E+02	1.4938E+02	1.8153E+02	1.3896E+02	1.6744E+02
Leukemia	STD	7.8564E+01	6.9273E+01	5.2495E+01	6.4638E+01	7.6812E+01	2.9753E+01
	AVG	<b>1.1529E+02</b>	1.8749E+02	1.6573E+02	1.7385E+02	1.9125E+02	1.3258E+02
Leukemia_GSE	STD	2.4826E+01	4.6394E+01	6.5927E+01	4.8563E+01	5.2841E+01	4.5294E+01
9476	AVG	<b>1.2038E+02</b>	1.9565E+02	1.6829E+02	1.8254E+02	1.4482E+02	1.7326E+02
Liver_GSE1452	STD	3.1739E+01	7.3658E+01	4.9382E+01	6.1527E+01	3.8964E+01	5.7435E+01
0_U133_2	AVG	<b>1.1459E+02</b>	1.8873E+02	1.5638E+02	1.7529E+02	1.9265E+02	1.3147E+02
Lung_GSE7670	STD	6.9453E+01	5.0826E+01	7.2574E+01	5.6238E+01	4.4692E+01	6.9315E+01
	AVG	1.4743E+02	1.9486E+02	1.6254E+02	1.8029E+02	<b>1.4065E+02</b>	1.6938E+02
SRBCT	STD	3.5287E+01	6.7493E+01	5.5729E+01	6.3846E+01	7.1254E+01	3.2576E+01
	AVG	<b>1.1385E+02</b>	1.8957E+02	1.5529E+02	1.7463E+02	1.9375E+02	1.3360E+02
Liver_GSE7642	STD	6.6594E+01	4.2738E+01	6.8463E+01	5.0295E+01	5.4629E+01	4.7843E+01
7	AVG	1.5968E+02	1.9628E+02	1.6759E+02	1.8146E+02	<b>1.4376E+02</b>	1.7140E+02
	AVG	<b>1.11</b>	6.00	3.22	5.00	2.22	3.00
	Rank	<b>1</b>	6	4	5	2	3

## 6. 结论

本文针对原始AO算法应用于高维医学特征选择任务时极易出现全局勘探与局部开发失衡,进而引发特征筛选效率低下的问题,对其进行优化改进,提出融合差异协作策略与稳定生长策略的改进算法DOAO,通过两种策略优化算法迭代逻辑,丰富种群多样性的同时平衡种群全局探索与局部开发能力。为验证算法全局优化能力,基于CEC 2017基准测试函数集,将DOAO与十种元启发式算法进行性能对比,证实DOAO在全局优化中的稳定性。为挖掘算法在特征选择领域的应用价值,通过二进制转换函数,将DOAO转化为二进制版本bDOAO,并与KNN分类器结合,构建包裹式特征选择模型bDOAO-KNN,选取9个高维医学数据集开展验证实验。挑选平均适应度值、平均分类错误率、平均特征子集规模和平均运行耗时作为评价指标,将所提方法与五种同行算法进行对比,结果表明,DOAO能够在控制运行耗时的前提下,高效剔除冗余特征,同时大幅提升分类精度,在高维医学特征选择任务中具备一定的竞争力与应用前景。

## 参考文献

- [1] Sajed, S., Sanati, A., Garcia, J.E., Rostami, H., Keshavarz, A. and Teixeira, A. (2023) The Effectiveness of Deep Learning vs. Traditional Methods for Lung Disease Diagnosis Using Chest X-Ray Images: A Systematic Review. *Applied Soft Computing*, **147**, Article ID: 110817. <https://doi.org/10.1016/j.asoc.2023.110817>

- [2] Wang, X., He, Q., Jian, W., Meng, H., Zhang, B., Jin, H., *et al.* (2024) Hybrid Feature Ranking and Classifier Aggregation Based on Multi-Criteria Decision-Making. *Expert Systems with Applications*, **238**, Article ID: 122193. <https://doi.org/10.1016/j.eswa.2023.122193>
- [3] Madhusudhanan, B., Sumathi, P., Karpagam, N.S., Mahesh, A. and Suhi, P.A.P. (2019) An Hybrid Metaheuristic Approach for Efficient Feature Selection. *Cluster Computing*, **22**, 14541-14549. <https://doi.org/10.1007/s10586-018-2337-2>
- [4] Yuan, C., Zhao, D., Heidari, A.A., Liu, L., Chen, Y., Wu, Z., *et al.* (2024) Artemisinin Optimization Based on Malaria Therapy: Algorithm and Applications to Medical Image Segmentation. *Displays*, **84**, 102740. <https://doi.org/10.1016/j.displa.2024.102740>
- [5] Wang, A., An, N., Chen, G., Li, L. and Alterovitz, G. (2015) Accelerating Wrapper-Based Feature Selection with K-Nearest-Neighbor. *Knowledge-Based Systems*, **83**, 81-91. <https://doi.org/10.1016/j.knosys.2015.03.009>
- [6] Govindarajan, M. and Chandrasekaran, R. (2010) Evaluation of K-Nearest Neighbor Classifier Performance for Direct Marketing. *Expert Systems with Applications*, **37**, 253-258. <https://doi.org/10.1016/j.eswa.2009.04.055>
- [7] Duankhan, P., Sunat, K., Chiewchanwattana, S. and Nasa-ngium, P. (2024) The Differentiated Creative Search (DCS): Leveraging Differentiated Knowledge-Acquisition and Creative Realism to Address Complex Optimization Problems. *Expert Systems with Applications*, **252**, Article ID: 123734. <https://doi.org/10.1016/j.eswa.2024.123734>
- [8] Storm, R. and Price, K. (1997) Differential Evolution—A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization*, **11**, 341-359. <https://doi.org/10.1023/a:1008202821328>
- [9] Ghasemi, M., Zare, M., Trojovský, P., Rao, R.V., Trojovská, E. and Kandasamy, V. (2024) Optimization Based on the Smart Behavior of Plants with Its Engineering Applications: Ivy Algorithm. *Knowledge-Based Systems*, **295**, Article ID: 111850. <https://doi.org/10.1016/j.knosys.2024.111850>
- [10] Heidari, A.A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M. and Chen, H. (2019) Harris Hawks Optimization: Algorithm and Applications. *Future Generation Computer Systems*, **97**, 849-872. <https://doi.org/10.1016/j.future.2019.02.028>
- [11] Li, S., Chen, H., Wang, M., Heidari, A.A. and Mirjalili, S. (2020) Slime Mould Algorithm: A New Method for Stochastic Optimization. *Future Generation Computer Systems*, **111**, 300-323. <https://doi.org/10.1016/j.future.2020.03.055>
- [12] Su, H., Zhao, D., Heidari, A.A., Liu, L., Zhang, X., Mafarja, M., *et al.* (2023) RIME: A Physics-Based Optimization. *Neurocomputing*, **532**, 183-214. <https://doi.org/10.1016/j.neucom.2023.02.010>
- [13] Abd Elaziz, M., Oliva, D. and Xiong, S. (2017) An Improved Opposition-Based Sine Cosine Algorithm for Global Optimization. *Expert Systems with Applications*, **90**, 484-500. <https://doi.org/10.1016/j.eswa.2017.07.043>
- [14] 张悦, 孙惠香, 魏政磊, 等. 具有自适应调整策略的混沌灰狼优化算法[J]. 计算机科学, 2017, 44(S2): 119-122, 159.
- [15] Liang, J.J., Qin, A.K., Suganthan, P.N. and Baskar, S. (2006) Comprehensive Learning Particle Swarm Optimizer for Global Optimization of Multimodal Functions. *IEEE Transactions on Evolutionary Computation*, **10**, 281-295. <https://doi.org/10.1109/tevc.2005.857610>
- [16] Seyyedabbasi, A. and Kiani, F. (2019) I-GWO and Ex-GWO: Improved Algorithms of the Grey Wolf Optimizer to Solve Global Optimization Problems. *Engineering with Computers*, **37**, 509-532. <https://doi.org/10.1007/s00366-019-00837-7>
- [17] Liang, H., Liu, Y., Shen, Y., Li, F. and Man, Y. (2018) A Hybrid Bat Algorithm for Economic Dispatch with Random Wind Power. *IEEE Transactions on Power Systems*, **33**, 5052-5061. <https://doi.org/10.1109/tpwrs.2018.2812711>
- [18] Chen, W., Zhang, J., Lin, Y., Chen, N., Zhan, Z., Chung, H.S., *et al.* (2013) Particle Swarm Optimization with an Aging Leader and Challengers. *IEEE Transactions on Evolutionary Computation*, **17**, 241-258. <https://doi.org/10.1109/tevc.2011.2173577>
- [19] Cervante, L., Xue, B., Zhang, M. and Lin Shang, (2012) Binary Particle Swarm Optimisation for Feature Selection: A Filter Based Approach. 2012 *IEEE Congress on Evolutionary Computation*, Brisbane, 10-15 June 2012, 1-8. <https://doi.org/10.1109/cec.2012.6256452>
- [20] Mirjalili, S., Mirjalili, S.M. and Yang, X. (2014) Binary Bat Algorithm. *Neural Computing and Applications*, **25**, 663-681. <https://doi.org/10.1007/s00521-013-1525-5>
- [21] Mafarja, M. and Mirjalili, S. (2018) Whale Optimization Approaches for Wrapper Feature Selection. *Applied Soft Computing*, **62**, 441-453. <https://doi.org/10.1016/j.asoc.2017.11.006>
- [22] Taghian, S. and Nadimi-Shahraki, M.H. (2019) Binary Sine Cosine Algorithms for Feature Selection from Medical Data. arXiv: 1911.07805.