Identification of Pulmonary Hypertension Animal Models Using a New Evolutionary Machine Learning Framework Based on Blood Routine Indicators

Jiao Hu1 · Shushu Lv2 · Tao Zhou3 · Huiling Chen1* · Lei Xiao1 · Xiaoying Huang4 · Liangxing Wang4 · Peiliang Wu4

Received: 5 June 2022 / Revised: 17 October 2022 / Accepted: 19 October 2022 / Published online: 28 November 2022
© Jilin University 2022

Abstract
Pulmonary Hypertension (PH) is a global health problem that affects about 1% of the global population. Animal models of PH play a vital role in unraveling the pathophysiological mechanisms of the disease. The present study proposes a Kernel Extreme Learning Machine (KELM) model based on an improved Whale Optimization Algorithm (WOA) for predicting PH mouse models. The experimental results showed that the selected blood indicators, including Haemoglobin (HGB), Hema
tocrit (HCT), Mean, Platelet Volume (MPV), Platelet distribution width (PDW), and Platelet–Large Cell Ratio (P-LCR), were essential for identifying PH mouse models using the feature selection method proposed in this paper. Remarkably, the method achieved 100.0% accuracy and 100.0% specificity in classification, demonstrating that our method has great potential to be used for evaluating and identifying mouse PH models.

Keywords Feature selection · Pulmonary hypertension · Whale optimization algorithm · Extreme learning machine

1 Introduction
The pulmonary artery is a major blood vessel that runs from the heart to the lungs. As a global health problem, Pulmonary Hypertension (PH) is estimated to affect about 1% of the global population[1]. PH is best defined by the concomitant presence of mean pulmonary arterial pressure (mPAP) > 20 mmHg, pulmonary arterial wedge pressure (PAWP) ≤ 15 mmHg, and pulmonary vascular resistance (PVR) ≥ 3 Wood, emphasizing the need for right heart catheterization with mandatory measurement of cardiac output and accurate measurement of PAWP[2]. Affecting people of all ages and all races, PH can be divided into five categories: pulmonary arterial hypertension; PH due to heart disease; PH due to lung disease; PH due to blockage of blood vessels, and PH due to other causes[3].

Transthoracic echocardiography is the most frequently employed non-invasive approach for evaluating PH. The
degree of tricuspid regurgitation, myocardial performance index, presence of pericardial effusion, pulmonary vascular resistance, cardiac index, and right atrial pressure are all echocardiographic characteristics that can be employed to forecast a patient’s survival with PH [4]. However, right-sided heart catheterization is relatively more accurate for diagnosing PH[5]. PH is a disease characterized by progressive remodeling of the distal pulmonary arteries, with all three layers of the arterial wall involved. In PH, typical arterial abnormalities include intimal and neointimal fibrosis, adventitial fibrosis with varying degrees of perivascular inflammation, and medial hyperplasia of pulmonary artery smooth muscle cells, adventitial fibrosis with varying degrees of perivascular inflammation. The pathological hallmark of severe PH is complex plexiform lesions consisting of endothelial over expansion, some of which display cancer-like features of monoclonal proliferation [5, 6]. In patients with PH, pulmonary vascular resistance can lead to Right Ventricle (RV) dilation, RV dysfunction, and RV failure, resulting in increased RV wall stress[7]. To maintain right ventricular output, the RV can accommodate a slow increase in pressure load by increasing contractility and wall thickness. Although the RV dilates to maintain stroke volume when pulmonary vascular resistance continues to rise, the RV cannot remodel indefinitely. The RV eventually decouples from the pressure load, resulting in RV failure[8].

PH is usually progressive without intervention, leading to right heart failure and death[6]. PH may impair heart function, resulting in breathing difficulties with effort, fatigue, chest discomfort, swelling of the legs or abdomen, and fainting or dizzy feeling [3, 5]. Therefore, PH requires aggressive treatment. Currently, the recognized basic measures and routine PH treatment include exercise rehabilitation, oxygen therapy, anticoagulation, calcium channel blockers, diuretics, and early electric cardioversion. Pregnancy should generally be avoided, and early termination is recommended for patients with PH. Genetic counseling and psychosocial support should also be considered [6]. In patients with PH, restricted exercise is necessary. Patients should be encouraged to exercise, but only when physical conditions allow. Recent randomized controlled trials have shown that supervised rehabilitation improves exercise capacity and quality of life[9]. Excitingly, some new targeted therapy methods have also been recently proposed. Chen and his colleagues discovered that nicotinamide phosphoribosyl-transferase contributes to in pulmonary vascular remodeling, and that inhibiting it might be a viable therapy for PH [10].

Despite improved survival of patients with PH in recent years, PH still cannot be cured, and its mechanisms have not been fully elucidated [11, 12]. In the research of PH, establishing a mouse model of PH is essential. Right heart catheterization is the most accurate measure of whether PH is established in mice. However, right heart catheterization, an invasive examination that requires anesthesia and intubation of the mouse, is a blind intubation method that can easily lead to failure of pressure measurement, bleeding, and even death of the mouse. Therefore, it is crucial to determine whether a model is established non-invasively. In recent years, machine learning methods have become increasingly widely used in the medical field. Predictive models based on machine learning can be used for medical decision-making and resource allocation and help medical personnel make medical model predictions [13, 14]. In this study, we explored whether a machine learning method could be created to test the successful establishment of a mouse model of PH.

Machine learning methods have been used for the medical diagnosis of a wide range of diseases. For example, Polat et al. [15] developed a support vector machine classification algorithm for diagnosing chronic kidney disease that uses a greedy search-based classifier and a best-first search-based wrapper to evaluate a subset. Abbad et al. [16] proposed using machine learning algorithms to detect and diagnose thyroid disorders that adopt efficient classifiers. Pashaei et al. [17] proposed a chimpanzee optimization-based feature selection method for biomedical data classification wrapper. Alsaeedi et al. [18] introduced a feature selection method based on the new cal- edonian crow learning algorithm to identify whether a population is infected with COVID-19. Lamba et al. [19] proposed a hybrid speech signal-based Parkinson’s disease diagnosis system for early diagnosing Parkinson’s disease by combining several feature selection methods and classification algorithms. Hu et al. [20] used a machine learning method based on an improved binary mutation quantum grey wolf optimizer combined with fuzzy k-nearest neighbor to predict the trend of serum albumin levels. Faisal et al. [21] proposed an ad hoc feature selection method to distinguish between Alzheimer’s disease, MCI and health control patients to reduce the model complexity considered when using machine learning methods. Hu et al. [22] proposed a prediction framework based on a kernel extreme learning machine combined with improved binary Harris hawk optimization to classify the severity of COVID-19. This paper combines optimization algorithms and machine learning methods to diagnose PH.

The remaining sections of this paper are organized as follows: Sect. 2 focuses on extracting relevant data sets. The methods used for feature selection are described in Sect. 3. In Sect. 4, the experimental results are analyzed. The final section discusses the experimental results.

2 Materials and Methods

2.1 Laboratory Animals

107 healthy specific pathogen-free male C57BL/6 mice aged 12–14 weeks, weighing 20–25 g (animal certificate number: A563).
SYXK (zhe) 2020–0014), were purchased from the Academy of Medical Sciences of Zhejiang Province and reared at the Experimental Animal Center of Wenzhou Medical University. All animal experiments were carried out following the Institutional Animal Care guidelines and were approved ethically by the Administration Committee of Experimental Animals, Laboratory Animal Center, Wenzhou Medical University.

2.2 Reagents and Instruments

Animal oxygen chamber (Changsha Huaxi Electronic Technology Co., Ltd., Hunan, China), Power Lab4/30 multi-channel physiological instrument (ML866, Australia Ed Instruments International Trading Co., Ltd.), powerlab multi-channel bio-signal recording system, pressure transducer probe and tee, 10 mL and 1 mL syringes, 0.9% saline, heparin sodium saline (10 U/mL), 20% urethane, III-0 silk thread, NIKON SMZ745T dissecting microscope, blood cell analyzer, mouse fixation plate, medical Tape, straight forceps, curved forceps, ophthalmic scissors, arterial clips, medical cotton balls, etc.

2.3 Methods

A total of 107 mice were randomly divided into the normoxic control group (50 mice) and the hypoxic pulmonary hypertension (HPH) group (57 mice). Mice in the HPH group were kept in an oxygen chamber with an oxygen concentration of (10 ± 1)% and a CO₂ concentration of (2 ± 1)% for 24 h and were reared for 21 days to induce PH. The mice in the normoxic control group were kept in an oxygen chamber with an oxygen concentration of 20.9% and a CO₂ concentration of 0.03% for 24 h and were raised for 24 h. Other conditions, such as food and water, were the same. After 21 days, 0.2–0.3 ml of blood was drawn from the abdominal aorta for blood routine measurements. A blood cell analyzer analyzed blood samples to obtain blood routine data, which were then used in machine learning experiments. The study flow chart is presented in Fig. 1. Table 1 lists the 25 blood routine indicators (features) measured in this study. At the same time, all the mice were connected to the multi-channel physiology instrument through the right heart catheterization to measure the right ventricular systolic pressure and the pressure of the pulmonary artery to ensure that the mice in the HPH group successfully became PH mice.

2.4 Statistical Analysis

Comparisons between the normoxic and HPH groups were evaluated using an independent sample \( t \) test. Measurement data are expressed as mean \((\bar{X}) \pm \text{standard deviation (SD)}\). \( P < 0.05 \) was considered a statistical significance. Analyses were performed using SPSS, version 21. Results are presented in Table 2.
3 Presented Method

This section introduces the related knowledge of the whale optimization algorithm and the rough set theory. Then, the whale-optimized gene selection algorithm based on a rough set is further elaborated.

3.1 Mathematical Model of WOA

Many new optimization algorithms have been proposed in recent years, such as Harris hawks optimization (HHO) [23], hunger games search (HGS) [24], colony predation algorithm (CPA) [25], Runge Kutta optimizer (RUN) [26], weighted mean of vectors (INFO) [27], and slime mould algorithm (SMA) [28]. There are many fields in which they have achieved remarkable success, such as optimization of machine learning model [29], scheduling problems [30–32], medical diagnosis [33, 34], fault diagnosis [35], solar cell parameter identification [36], multi-objective problems [37, 38], combination optimization problems [39], and global optimization [40, 41].

Apart from the above, the principles of the Whale Optimization Algorithm (WOA) are derived from the modeling of humpback whale hunting behavior. Humpback whale hunting methods can usually be summarized into two behaviors. The first is sprint feeding, where when a humpback whale finds prey, it will dash directly towards the prey and open its mouth to swallow it directly. The second way is to rise upwards in a spiral-like position at a distance of 15 m from the surface, spitting out bubbles of different sizes, as they arise in order for all the bubbles to reach the surface simultaneously. The humpback whales are surrounded by these bubbles and swallow the prey when they reach the surface. This section details the basic whale algorithm based on the mathematical models of these two predatory behaviors.

3.1.1 Spiral Position Update

The mathematical model for the spiral period is shown in Eqs. (1) and (2):

\[
X(t + 1) = D^l \cdot e^{hl} \cdot \cos(2\pi l) + \bar{X}^*(t),
\]  

Table 1  List of the features used in this study and their abbreviations

| Features                          | Abbreviation |
|----------------------------------|--------------|
| F1 White blood cell              | WBC          |
| F2 Red blood cell                | RBC          |
| F3 Haemoglobin                   | HGB          |
| F4 Mean corpuscular volume       | MCV          |
| F5 Blood platelet                | PLT          |
| F6 Neutrophil percentage         | NEU%         |
| F7 Lymphocyte percentage         | LYM%         |
| F8 Monocyte percentage           | MON%         |
| F9 Eosinophils percentage        | EOS%         |
| F10 Basophils percentage         | BAS%         |
| F11 Neutrophil count             | NEU          |
| F12 Lymphocyte count             | LYM          |
| F13 Monocyte count               | MON          |
| F14 Eosinophils count            | EOS          |
| F15 Basophils count              | BAS          |
| F16 Hematocrit                   | HCT          |
| F17 Mean corpuscular hemoglobin  | MCH          |
| F18 Mean corpuscular hemoglobin concentration | MCHC |
| F19 Red blood cell distribution width coefficient of variation | RDW-CV |
| F20 Red blood cell distribution width-size distribution | RDW-SD |
| F21 Mean platelet volume         | MPV          |
| F22 Plateletcrit                 | PCT          |
| F23 Platelet distribution width   | PDW          |
| F24 Platelet–large cell ratio    | P-LCR        |
| F25 Platelet–large cell count    | P-LCC        |
Table 2 Blood routine indicators in the normoxic group and HPH group

| Index | Normoxic group (n = 50) | HPH group (n = 57) | p value |
|-------|-------------------------|--------------------|---------|
| F1    | WBC 3.516±1.526         | 7.585±4.129        | 0.000   |
| F2    | RBC 7.271±0.998         | 8.865±0.51001      | 0.000   |
| F3    | HGB 124.140±18.732      | 164.930±12.068     | 0.000   |
| F4    | MCV 42.256±0.992        | 46.698±2.926       | 0.000   |
| F5    | PLT 280.260±114.212     | 267.280±75.172     | 0.496   |
| F6    | NEU% 22.176±7.006       | 34.379±5.036       | 0.000   |
| F7    | LYM% 62.362±9.353       | 49.416±5.726       | 0.000   |
| F8    | MON% 13.580±2.799       | 13.623±2.659       | 0.936   |
| F9    | EOS% 1.134±0.602        | 1.781±0.969        | 0.000   |
| F10   | BAS% 0.748±0.318        | 0.802±0.485        | 0.506   |
| F11   | NEU 2.207±1.081         | 2.542±1.317        | 0.000   |
| F12   | LYM 0.474±0.212         | 1.021±0.587        | 0.000   |
| F13   | MON 0.039±0.027         | 0.144±0.132        | 0.000   |
| F14   | BAS 0.026±0.014         | 0.066±0.065        | 0.000   |
| F15   | HCT 30.760±4.339        | 41.426±3.813       | 0.000   |
| F16   | MCH 16.990±7.474        | 18.591±4.83        | 0.000   |
| F17   | MCHC 402.200±17.458     | 399.460±23.769     | 0.494   |
| F18   | RDW-CV 8.552±0.272      | 9.593±1.417        | 0.000   |
| F19   | RDW-SD 24.314±0.575     | 26.263±1.029       | 0.000   |
| F20   | MPV 12.688±1.494        | 14.009±1.304       | 0.000   |
| F21   | PCT 0.3479±0.132        | 0.3741±0.116       | 0.278   |
| F22   | PDW 14.622±2.353        | 16.281±0.194       | 0.000   |
| F23   | PLCR 0.359±0.097        | 0.4743±0.077       | 0.000   |
| F24   | P-LCC 93.120±30.563     | 127.210±49.672     | 0.000   |

\[ \overline{D} = |X(t) - \overline{X}|, \]

where vector \( X(t) \) is the current position of the whale, vector \( \overline{X}(t + 1) \) is the position of the whale after iteration, and vector \( \overline{X}^*(t) \) is the best position. \( b \) is a constant that controls the shape of the logarithmic spiral. \( l \) is a random number between \([-1, 1]\). When \( l \) reaches 1, the whale is farthest from the optimal position; when \( l \) reaches -1, the whale is closest to the optimal position. \( \cdot \) denotes point multiplication.

3.1.2 Reduce the Encirclement

\[ \overline{X}(t + 1) = \overline{X}^*(t) - \overline{A} \cdot \overline{D}, \]

\[ \overline{D} = |\overline{C} \cdot \overline{X}^*(t) - \overline{X}(t)|, \]

\[ \overline{A} = 2a \cdot \overline{r}_1 - a, \]

\[ \overline{C} = 2\overline{r}_2, \]

\[ a = 2 - \frac{2t}{T_{\text{max}}}. \]

where \( \overline{r}_1, \overline{r}_2 \), are random numbers between (0, 1), vector \( \overline{A} \) and vector \( \overline{C} \) are used to control the position of the whale in the process of position update. \( a \) is an important variable in the WOA, because \( a \) affects vector \( \overline{A} \) by indirectly controlling the mode of travel of whale. As shown in Eq. (7), a linearly decreases from 2 to 0. \( T_{\text{max}} \) is the maximum number of iterations.

The above sections combined are the previously described bubble net attacks. A random number is used to select the method of the bubble net attack, and the probability of this random number is usually set to \( p \). The probability of \( p \) is usually set to 0.5.

Depending on the parameter \( p \), the bubble net attack can be made in different ways. They are as follows:

\[ \overline{X}(t + 1) = \begin{cases} \overline{X}(t) - \overline{A} \cdot \overline{D}, & \text{rand} < p \\ \overline{D} \cdot e^b \cdot \cos(2\pi l) + \overline{X}(t), & \text{rand} \geq p \end{cases} \]

3.1.3 Random Search

The random search process evolves into a mathematical model as shown in Eq. (9):

\[ \overline{X}(t + 1) = \overline{X}_{\text{rand}} - \overline{A} \cdot \overline{D}_{\text{rand}} \]

\[ \overline{D}_{\text{rand}} = |\overline{C} \cdot \overline{X}_{\text{rand}} - \overline{X}(t)| \]

\( \overline{X}_{\text{rand}} \) is a randomly selected individual from a population of whales.

Which to use, the bubble spiral attack way or the random search way, is determined by the vector \( \overline{A} \) of Eq. (3.5). When \( \overline{A} \geq 1 \), the position of the WOA is updated in the random search way, and when \( \overline{A} < 1 \), the position of the WOA is updated in the bubble spiral attack way.

The simplified pseudo-code of WOA is listed in Algorithm 1.
Algorithm 1: Pseudo-code of WOA

Begin

Initialize the parameters: $T_{\text{max}}$, Population size $N$, $p$;

Initialize the whale population $X$ randomly;

While $t < T_{\text{max}}$

Calculate the fitness for each individual $X_i$ in the whale population;

Update $X^*$ and the best fitness;

Calculate $\alpha$ according to Eq. (7);

Calculate vectors $\vec{A}$ and $\vec{C}$ according to Eq. (5), Eq. (6);

For $i = 1, 2, \ldots, N$ (each search agent)

If $| \vec{A} | \geq 1$

Update population according to Eq. (9);

Else

$r = \text{rand} ()$;

If $r < p$

Update population according to Eq. (3);

Else

Update population according to Eq. (1);

End

End For

$t = t + 1$;

End While

Return the best fitness and $X^*$ as the best solution;

End

3.2 Whale Optimization Algorithm with Hybrid Conversion Mechanism (HCWOA)

In this section, a hybrid conversion mechanism inspired by TRIZ (Teoriya Resheniya Izobretatel’skih Zadatch) creative solution is introduced and combined with the basic WOA to maintain population diversity in the search process to explore more efficient gene interactions, resulting in new individuals.

TRIZ [42] is a knowledge-based, human-oriented systematic methodology for inventive problem solving. The TRIZ theory itself is based on the practice of decomposing systems into subsystems, distinguishing between useful and harmful functions, and these decompositions are problem and context-dependent and inherently stochastic. Furthermore, creatively solving problems is central to the innovation process. Although there are several theories and methods, the standard procedure for dealing with them is to use randomized trial and error. The findings of TRIZ and evolutionary algorithms support the idea that creativity can be systematically understood and developed.

One of the TRIZ tools is the 40 invention principles, which consist of a set of generic solutions that solve technical contradictions in many fields. The principles are organized according to the contradictions they solve, which makes it easy to deal with problems. In recent years, the TRIZ theory has been used to solve different problems in various industries. For example, Paolo et al. [43] proposed an innovation management framework to rely on partners with TRIZ skills to coordinate customized innovation processes. Vladimir et al. [44] analyzed several fashion inventions and apparel production techniques from a TRIZ perspective. Liu et al. [45] explored the impact of TRIZ learning on graduate students and showed that participants with TRIZ learning experience produced more novel design solutions and demonstrated the positive impact of TRIZ learning on bio-inspired designs. Khadija et al. [46] used a TRIZ contradiction matrix to address the main issues that arose during TRIZ matrix development. Christian et al. [47] improved the sustainability of different devices to a great extent by reducing environmental impact through TRIZ strategies.

In this paper, a Hybrid Conversion (HC) mechanism derived from TRIZ theory is used to help expand the advantages of each stage of WOA by addressing the shortcomings of the three different stages: spiral position update, narrowing envelope, and random search. The first operation of HC is crossover behavior. In whale populations, individuals with little difference in fitness tend to cluster together, and various types of crossover variation will be in this small population as the population evolves. However, this situation can make the population evolve to a limited extent and easily fall into local optima in the global optimization of complex problems. To alleviate this situation, we propose a crossover behavior. First, the whole whale population is divided into K groups based on time fitness values, and then individuals in each group are fused with individuals in other groups to form a new population. The specific process is shown in Fig. 2.

The second operation in HC is local optimization, as shown in Fig. 3. This operation is divided into three main operators: dynamic partitioning, mutation, and transition. These three operators are described in detail below.

3.2.1 Dynamic Partitioning Operator

In this operator, the optimal individual in the population is invariant, and the other individuals are decomposed into smaller blocks. The specific size $S$ of the block is generated dynamically by the algorithm. If the individual dimension is not divisible by $S$, then the last piece is used as a reminder of the size of the block. In particular, if the remainder is 1, then it is merged with the previous block. This setting is mainly used to evaluate individual dimensions to find the more important features for classification during the feature selection process.
Fig. 2 Schematic illustration of crossover operation

Fig. 3 Schematic illustration of individual changes
3.2.2 Mutation Operator

This operator is roughly the same as the commonly used mutation operator. The specific operation is to subdivide each individual block into two groups. One of the two groups will remain unchanged, and the other will perform a mutation operation. The specific mutation is to generate a random number between 0 and 1 randomly. If this random number is less than 0.5, a Gaussian mutation operation is performed, and if it is greater than 0.5, a Cauchy mutation operation is performed, and then the two groups are reunited into one block. The mutation operation during the feature selection operation means that the original 0 is mutated to 1, and 1 is mutated to 0. For example, 0010101 to 1101010.

3.2.3 Transition Operator

This is an inter-block operation. In this operator, two random block indices are randomly generated with random numbers between 0 and K. K indicates the total number of blocks each individual is divided into. For the two selected blocks, the positions of the two blocks are swapped in the individual. The blocks are then re-linked into a new individual.

3.3 Classification Based on KELM

Extreme Learning Machine (ELM) is a special variant of the fast single hidden layer feedforward neural algorithm. Huang et al. [48] introduced regularization schemes and kernel functions into ELM to obtain the kernel extreme learning machine (KELM). KELM can improve the prediction performance of the model while retaining the advantages of ELM. Among them, ELM is a single hidden layer feedforward neural network whose learning objective function $F(x)$ can be represented by the matrix:

$$F(x) = h(x) \times \beta = H \times \beta = L,$$

where $x$ is the input vector, $h(x)$ and $H$ are the output of the hidden layer node, $\beta$ is the output weight, and $L$ is the desired output.

Network training can be turned into a problem solved by a linear system. $\beta$ can be determined from $\beta = H^\ast \cdot L$, where $H^\ast$ is the generalized inverse matrix of $H$. In addition, to enhance the stability of the neural network, the regularization coefficient $C$ and the unit matrix $I$ are introduced, and the least-squares solution of the output weights is as follows:

$$\beta = H^T \left( HH^T + \frac{I}{C} \right)^{-1} L.$$  

The kernel function is introduced into ELM, and the kernel matrix is

$$\Omega_{ELM} = HH^T = h(x_i)h(x_j) = K(x_i, x_j),$$

where $x_i$ and $x_j$ are the test input vectors, then Eq. (3.11) can be expressed as

$$F(x) = [K(x_1, x_1); \ldots ; K(x_n, x_n)] \left( \frac{I}{C} + \Omega_{ELM} \right)^{-1} L,$$

where $(x_1, x_2, \ldots, x_n)$ is the given training sample, $n$ is the number of samples, and $K()$ is the kernel function.

In this paper, the routine blood data of pulmonary hypertension were classified, and the training set and test set were generated by tenfold cross-validation. The regularization coefficient $C$ and the kernel function parameter $S$ were respectively, obtained after HCWOA optimization, and the kernel function was selected as the RBF Gaussian kernel function.

To better classify pulmonary hypertension blood routine data, HCWOA and KELM are combined for feature selection. The method can provide interpretable results for classification and help physicians in medical diagnosis. The flow chart of the method in this paper is as follows and the flowchart is shown in Fig. 4.

Step 1: Initialize the input parameters of HCWOA, including the population size, the boundary of the search space, the maximum number of iterations, and the dynamic block $S$.

Step 2: Initialize a random binary whale population.

Step 3: Obtain a subset of features (1: features are selected, 0: features are not selected) according to the whales’ locations.

Step 4: Use the selected feature subset to calculate the fitness of each whale as follows:

$$\text{Fitness} = \alpha \cdot E + \beta \cdot \frac{|R|}{|d|},$$

where $E$ denotes the classification error rate of KELM, $|d|$ denotes the number of selected feature subsets, and $R$ denotes the total number of selected features. In addition, $\alpha$ and $\beta$ are two weights that measure the importance of the classification error rate and the size of the selected feature subset. we set $\alpha = 0.99$ and $\beta = 1 - \alpha$.

Step 5: Select the whale with the smallest fitness as the optimal position.

Step 6: Update the control parameters $a$, $A$ and $C$ according to Eqs. (5–7)

Step 7: Update the whale population location.

Step 8: Perform the hybrid conversion mechanism for the whales other than the optimal position to obtain a new population.
Step 9: Select the optimal one to reconstitute the new population according to the greedy idea.

Step 10: Judge whether the end condition is reached, if not, repeat steps 4–9.

Step 11: Return to the optimal position.

4 Experiment and Results

4.1 Validation of Function Optimization

In this paper, to test the performance of the proposed HCWOA, 28 benchmark functions were used as test sets, including 4 unimodal functions (F1–F4), 6 multimodal functions (F5–F10), 4 fixed modal functions (F10–F15), 6 hybrid functions (F15–F20) and 8 composition functions (F21–F28). Table 3 lists the descriptions of the 28 benchmark functions. These benchmark functions represent a variety of the most complex mathematical optimization problems. Therefore, they are often used to evaluate the comprehensive ability of algorithms. The average performance of all compared algorithms was further compared statistically using the Freidman test, and the average rank was given according to comparison results.

The comparison algorithm adopted includes original WOA, common meta-heuristic algorithm: Bat Algorithm (BA) [49], Firefly Algorithm (FA), Moth-Flame Optimization (MFO) [50], Differential Evolution (DE) [51], and improved advanced algorithm: differential Evolution algorithm based on Chaotic Local Search (DECLS) [52], Chaotic and Gaussian Particle Swarm Optimization (CGPSO) [53], Particle Swarm Optimization with Aging Leader and Challengers (ALCPSO) [54].

These tests were conducted out using a Microsoft Windows Server 2012 R2 data center version of windows with 128 GB RAM and an Intel (R) Xeon (R) E5-2650 v4 (2.20 GHz) CPU, with code written in MATLAB R2014b.
All techniques were evaluated under the identical conditions to ensure a fair assessment. The population size was set to 30 and the maximum number of evaluations was set to 1500. To reduce other effects, all algorithms were tested on the benchmark functions for 30 times. In this paper, the numerical results of these methods were selected based on the average (Avg.) and standard deviation (Std.) of the optimal function values. Avg. is used to evaluate the global search ability, and Std. is used to evaluate the algorithm’s robustness. In addition, the optimal results for each problem are marked in bold to display the optimal results clearly. The Wilcoxon signed-rank test [55], a nonparametric statistical test at the 0.05 significance level, was used to determine whether the improvement was statistically significant. The symbols “+/−” indicate that the proposed method is better than, equal to and worse than the other competitors, respectively.

Table 4 shows the optimization results of HCWOA and 8 competitors on 28 benchmark functions. The bolded values in the table indicate the best optimization on that test function. It can be seen that HCWOA achieves optimal results on 24 functions. In particular, HCWOA can directly find the optimal values on the unimodal functions. On the multimodal functions (F5–F10), HCWOA reaches the optimum on some functions. Although HCWOA does not optimize as well as FA and CGPSO on F5, F9 and F10, there is only a small difference between the results obtained by HCWOA
| Item | HCWOA Avg | WOA Avg | BA Avg | FA Avg | MFO Avg | DE Avg | DECLS Avg | CGPSO Avg | ALCPSO Avg |
|------|------------|---------|--------|--------|---------|--------|-----------|-----------|------------|
| F1   | 0.00E+00   | 1.37E-74| 2.18E+00| 1.52E+04| 1.69E+03| 4.58E-04| 1.41E-03 | 1.03E-03 | 1.09E-05   |
| F2   | 0.00E+00   | 4.98E-74| 1.01E+00| 1.50E+03| 5.30E+03| 2.25E-04| 2.71E-03 | 2.32E-03 | 1.27E-05   |
| F3   | 0.00E+00   | 9.34E-51| 8.05E+01| 6.39E+01| 3.19E+01| 2.10E-03| 1.16E-02 | 1.22E-02 | 2.87E-03   |
| F4   | 0.00E+00   | 2.89E-50| 3.55E+02| 7.97E+00| 1.98E+01| 4.53E-04| 1.38E-02 | 1.21E-02 | 1.06E-02   |
| F5   | 0.00E+00   | 4.56E+04| 2.62E+04| 2.76E+04| 1.97E+04| 3.11E+04| 1.62E+01 | 4.25E+00 | 2.62E+03   |
| F6   | 0.00E+00   | 1.54E+04| 1.30E+01| 3.71E+03| 1.22E+04| 5.06E+03| 2.16E-01 | 4.66E+00 | 1.18E+03   |
| F7   | 0.00E+00   | 4.43E+01| 1.12E+01| 4.83E+01| 6.91E+01| 1.32E+01| 4.02E-03 | 4.17E+01 | 1.29E+01   |
| F8   | 0.00E+00   | 2.86E+01| 5.87E+00| 1.94E+00| 8.18E-00| 1.74E+00| 3.59E-03 | 4.75E-03 | 3.18E+00   |
| F9   | 0.00E+00   | -1.07E+04| -1.10E+04| -7.42E+03| -3.55E+03| -8.73E+03| -9.76E+03| -1.26E+04| -2.30E+04   |
| F10  | 1.43E+03   | 1.57E+03| 7.06E+02| 3.17E+02| 9.73E+02| 5.03E+02| 7.90E-04 | 3.89E+03 | 6.32E+02   |
| F11  | 0.00E+00   | 1.89E-15| 2.85E+02| 2.60E+02| 1.70E+02| 8.78E+01| 3.13E-04 | 2.05E-04 | 6.79E+01   |
| F12  | 0.00E+00   | 1.04E-14| 3.04E+01| 1.53E+01| 3.61E+01| 8.36E+00| 7.60E-04 | 3.38E+04 | 2.17E+01   |
| F13  | 6.45E-02   | 5.30E-01| 7.45E-01| 3.60E+07| 9.29E+00| 2.84E-04| 1.43E-04 | 5.48E+05 | 4.48E-02   |
| F14  | 0.00E+00   | 3.24E-15| 3.00E+00| 4.43E-01| 5.58E+00| 1.21E-03| 4.68E-03 | 3.01E+03 | 9.48E-01   |
| F15  | 0.00E+00   | 4.04E-03| 2.98E+00| 1.35E+02| 9.01E+00| 3.26E-03| 1.50E-03 | 2.59E-02 | 3.02E+00   |
| F16  | 0.00E+00   | 5.04E-03| 6.04E+00| 1.59E+01| 4.36E+01| 8.28E-03| 3.28E-03 | 4.15E-02 | 3.02E+00   |
| F17  | 0.00E+00   | 2.21E+01| 7.67E+02| 8.24E+00| 4.92E+00| 8.66E-06| 4.61E+06 | 5.33E-01 | 3.02E+00   |
| F18  | 0.00E+00   | 1.25E+01| 7.78E+00| 6.36E+06| 2.75E-05| 2.09E-05| 1.01E-05 | 4.83E-01 | 3.02E+00   |
| F19  | 6.45E-02   | 5.30E-01| 7.45E-01| 3.60E+07| 9.29E+00| 2.84E-04| 1.43E-04 | 5.48E+05 | 4.48E-02   |
| F20  | 3.63E-02   | 3.19E+00| 2.31E+00| 1.30E+00| 3.54E+00| 9.62E-01| 7.28E-05 | 9.35E-05 | 3.74E+00   |
| F21  | 4.43E-04   | 3.71E+05| 9.42E+04| 1.34E+00| 2.20E+05| 7.64E+04| 2.72E+05 | 1.17E+05 | 4.47E+04   |
| F22  | 3.26E-02   | 5.04E-03| 2.06E+00| 1.07E+00| 3.40E+00| 1.22E+00| 7.30E-05 | 5.05E-05 | 3.67E+00   |
| F23  | 2.39E+00   | 8.16E+06| 6.15E+07| 3.94E+09| 5.47E+08| 3.69E+07| 1.24E+08 | 1.62E+08 | 4.14E+08   |
| F24  | 5.53E+03   | 4.34E+07| 3.71E+05| 9.20E+08| 5.34E+07| 1.95E+03| 1.02E+04 | 6.98E+06 | 3.32E+06   |
| F25  | 2.90E+03   | 4.10E+03| 3.98E+03| 4.16E+03| 3.60E+03| 3.48E+03| 2.99E+03 | 3.10E+03 | 3.87E+03   |

Table 4 Comparison of optimal values between HCWOA and famous MAs
and the minimum values. In addition, for the fixed modal functions, although HCWOA's optimization result on F17 is worse than BA, HCWOA’s optimization results are optimal on the other functions. HCWOA always has the best optimization results for other test functions. This is because the operation for populations and individuals in HCWOA can help avoid local optima.
Table 5 shows the $p$ value of the Wilcoxon signed-rank test. $p$ value less than 0.05 indicates statistical significance. Values less than 0.05 are bolded in the table, indicating that HCWOA is significantly better than the comparison method. As can be seen from the table, HCWOA significantly outperforms BA, FA, and DE on the 28 benchmark functions. Although on F5–F6 and F11–F14, the optimization results of HCWOA and WOA are not significantly different, it can be seen that more than 95% of the data in the table are less than 0.05. Therefore, it can be said that HCWOA has better optimization results compared with the other 8 algorithms.

Figure 5 shows the convergence effect of HCWOA on unimodal, multimodal, fixed modal, hybrid, and composition functions. The convergence curves were obtained from the evaluation process by selecting 50 points in sequence and plotting them as smooth curves. It can be seen from the figure that HCWOA has not only the smallest convergence but also the fastest convergence on the unimodal and multimodal functions. On the fixed modal function, although the convergence speed of HCWOA is not the fastest, the convergence value is the smallest. For F21 and F27, although the convergence speed of HCWOA is similar to those of DECLS and CGPSO, the convergence value of HCWOA is the smallest. As can be seen from the convergence figure, the population diversity and convergence accuracy of HCWOA are greatly improved compared to the basic WOA.
4.2 Feature Selection Experiment in the PH Data Set

To verify the effectiveness of the proposed HCWOA–KELM method, bHCWOA–KELM was compared with a range of machine learning methods, including BP, CART, Random Forest, AdaBoostM1, ELM and KELM traditional classification algorithms. Among them, the BP algorithm, CART, RandomF and AdaBoostM1 used self-built classifiers in MATLAB. The ELM algorithm can be found at http://www.ntu.edu.sg/eee/ics/cv/egbhuang.htm. KELM is based on ELM with the addition of kernel functions. For a fair comparison, all experiments were performed in the same simulated environment. The number of hidden neurons was 10 for the BP algorithm and 20 for the ELM algorithm. The regularization factor C and the kernel function parameter S for KELM were set to 2 and 4, respectively. The settings for the remaining classifiers were specified by convention. To get a fair and impartial outcome, the classification performance was assessed employing tenfold cross-validation (CV) methodology. In addition, to evaluate the performance of bHCWOA–KELM, we used four popular metrics, i.e., specificity, sensitivity, classification accuracy (ACC), and MCC.

The classifier’s effectiveness is verified using four mutual rules based on the confusion matrix. The complete definitions of these metrics are provided in [56, 57]. Here, we present their formulations to avoid discussions beyond the scope of this study:

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN},
\]

\[
\text{Specificity} = \frac{TN}{FP + TN},
\]

\[
\text{Sensitivity} = \frac{TP}{TP + FN},
\]

\[
\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}.
\]

The specific experimental results are as follows. To verify its effectiveness in feature selection, the proposed bHCWOA–KELM algorithm was compared with commonly used classification methods without feature selection. The comparison results of the seven classifiers are shown in Fig. 6. It can be seen from the figure that, although the differences between bHCWOA–KELM and AdaBoost in terms of specificity and classification accuracy are small, bHCWOA–KELM is much better than the other algorithms in terms of sensitivity and MCC. The experimental results show that bHCWOA–KELM can handle the classification of high-pressure pulmonary artery blood routine data, suggesting that it can provide some help for experts and doctors in real life.

From the above, it is clear that bHCWOA–KELM is significantly more effective in feature selection than the common classification methods. To further evaluate the effectiveness of this method on the routine high-pressure pulmonary artery blood data set, bHCWOA–KELM was compared with the commonly used bMFO, BPSO, BSSA, BBA, and bWOA in combination with KELM. The convergence curves of these six algorithms are shown in Fig. 7. The convergence values are the fitness set in Sect. 3 of this paper. As shown in the figure, the effectiveness of this algorithm

![Fig. 6 Comparison effect of bHCWOA on 6 classifiers](image1)

![Fig. 7 Convergence evolution trends of the six methods](image2)
is measured in terms of both convergence value and convergence speed. From the figure, it can be seen that bHCWOA–KELM has not only the fastest convergence speed but also the smallest convergence value. These results demonstrate that bHCWOA–KELM can effectively establish a proper balance between classification accuracy and feature subset size when dealing with the feature selection problem.

Finally, to further help physicians in the analysis of the classification results, the specific selected features from the tenfold cross-validation results were analyzed. The specifically selected features in each experiment are shown in Fig. 8. As can be seen from the figure, the most frequently selected features are HGB, HCT, MPV, PDW and p-LCR. Further analysis shows that these five features are the most important for the final classification results. Therefore, bHCWOA–KELM can provide accurate classification results and help physicians analyze which features are the key factors in routine blood data.

5 Discussion

5.1 The Analysis Method and Parameter Selection

The optimization properties of the HC operator based on TRIZ theory, HCWOA can obtain better global optimal solutions with better feature selection ability, as shown by the experimental results. The dynamic segmentation operator, variation, and transposition operations can effectively partition and optimize the structure of the search space. This can improve the performance of the whole population from the individual whales by gene exchange and global optimization. The three stages of WOA, namely, spiral position update, narrowing envelope, and random search, helped to expand each stage’s advantages by borrowing the hybrid transformation mechanism distilled from TRIZ theory. This is because TRIZ theory itself is based on the practice of decomposing systems into subsystems, distinguishing useful and harmful functions, and these decompositions depend on the problem and the environment and are somewhat stochastic in nature.

In addition, to verify the practical application capability of HCWOA, we apply its discretization and KELM in combination with the PH data set. From the experimental results, it can be seen that such a combination not only enhances the performance of KELM but also shows excellent performance in the feature selection problem. In addition, bHCWOA–KELM can accurately select the features in the PH data set that plays a deterministic role in the classification effect.

5.2 Physiological Significance

In this study, machine learning was used to identify whether or not a mouse model of PH was successfully established by analyzing blood samples. Several key features were identified, including HGB, HCT, MPV, PDW and P-LCR.

It is well-known that normal HGB is a tetramer composed of two α-like polypeptide subunits and two β-like polypeptide subunits[58]. HGB’s primary purpose is to carry oxygen from the lungs to peripheral tissues and carbon dioxide from peripheral tissues to the lungs for excretion, therefore, controlling the body’s blood oxygen balance[59]. Therefore, in an oxygen-deficient environment, HGB is bound to change. Anna Hauser et al. reported increased HGB content in hypoxic environments[60]. Steven Deem et al. study found that oxy-HGB rapidly oxidizes NO to nitrate and methemoglobin, resulting in a greatly weakened duration and magnitude of the vasodilation effect of NO in the pulmonary circulation, thus manifesting as increased pulmonary vascular
resistance [61, 62]. Meanwhile, cell-free hemoglobin as a pro-inflammatory oxidant [63] was found to cause not only acute lung injury when cell-free hemoglobin is present in the trachea of mice, but also airspace inflammation and alveolar capillary barrier damage [64, 65]. Based on these studies, it can be seen that hypoxia can lead to increased HGB, while elevated HGB damages alveoli and increases pulmonary vascular resistance. This study found that the HGB of mice in the HPH group was 1.33 times higher than that in the control group \((p = 0.000)\), suggesting that HGB may be a promising predictor for evaluating models of PH in mice.

HCT refers to the volume fraction of red blood cells in the blood and serves as one of the most important indicators of a patient’s blood status. The content of HCT is of great significance to diagnosing many diseases, such as inflammation, anemia, polycythemia, and blood stimulants [66]. The content of HCT depends on the level of erythropoietin (EPO) produced by the kidneys. The partial pressure of oxygen can regulate EPO. Under hypoxic conditions, EPO will increase, resulting in increased HCT [67]. In addition, many studies have shown that, within a certain range, the ability of the circulatory system to deliver oxygen increases with the elevation of HCT [68]. Numerous epidemiological data suggest that there is a close relationship between HCT and cardiovascular disease [69]. Similar to these studies, Stauffer E and Beall CM et al. found that in healthy individuals, prolonged hypoxic conditions lead to elevated HCT [70, 71]. In line with these findings, our research confirmed that the HCT level of mice in the HPH group was higher than that of the control group \((p = 0.000)\) (1.35 times), and the HCT level may be effective in assessing PH models in mice.

MPV is defined as the mean volume of platelets in the blood [72]. Under physiological conditions, MPV is inversely proportional to the number of platelets to maintain normal platelet function, which means that an increase in MPV will simultaneously lead to a decrease in platelets [73]. Studies have found that in the presence of hypoxia, the destruction of platelets is increased, and at the same time, MPV is increased [74]. MPV may also be a predictive marker of cardiovascular events, closely related to thrombosis and inflammation [75, 76]. Tromsø et al. found that increased MPV was a predictor of venous thromboembolism (VTE) [77]. The related meta-analysis found that the MPV of VTE patients was significantly higher than that of the control group [78]. Steiropoulos et al. found that the MPV of COPD patients was significantly higher than that of healthy people [79]. Numerous studies have also shown that severe hypertensive disease and related target organ damage are associated with elevated MPV [80–82]. Similar to the findings of their study, we also found that the MPV of mice in the HPH group was higher than that in the control group, about 1.10 times that of the control group. Therefore, MPV might have a significant predictive value in distinguishing mice with PH.

Similar to MPV, PDW is also a volume parameter, meaning the distribution of platelet size, which is the relative height of the 20% platelet size distribution histogram [83, 84]. Increased PDW means uneven platelet volume, suggesting platelet activation [85]. PDW was positively correlated with mean pulmonary artery pressure [86]. Some possible mechanisms are: activated platelets play a key role in pulmonary vascular remodeling and thrombosis [87, 88]. Sonali Jindal et al. also found a positive correlation between PDW and microvascular complications [89]. Multiple studies have shown that PDW in patients with idiopathic pulmonary hypertension is significantly higher than that in controls [90]. Consistent with these findings, our experimental results also showed that the PDW of mice in the HPH group was 1.11 times that of the control group \((p = 0.000)\).

P-LCR, an indicator used to quantify large platelets, is the proportion of platelets larger than 12 fl in the total number of platelets [91]. Large platelets are hypothesized to be more active [92, 93]. Numerous studies have shown that large platelets contain more particles and receptors and have higher hemostasis, thrombosis, and pro-inflammatory abilities than small platelets [94]. In relevant animal experiments, it was found that after the consumption of platelets, the newly generated platelets were significantly larger than the previous platelets [95]. The same is true after chemotherapy in humans [96]. In our experiment, we found that the P-LCR of mice in the HPH group was higher than that in the control group, indicating it could potentially serve as a valuable predictor \((p = 0.000)\).

However, because our experiments may have many limitations, these results require more rigorous interpretation and in-depth thinking. First, our experimental results partially depend on our experimental apparatus, resulting in possible deviations in the results. Second, the number of factors we selected was still too small. In further research, we should expand the research scope to blood gas analysis, coagulation mechanism, inflammatory factors, etc. Third, the sample size we studied was not large enough, and the next step should be to increase the sample size and improve the accuracy of the study. Based on the fact that the proposed algorithm has such an excellent performance, it can be applied to other fields as well, such as recommender system [97–100], human activity recognition [101], text clustering [102], medical image augmentation [103, 104], microgrids planning [105], named entity recognition [106], COVID-19 classification [107–109], and object tracking [110].
6 Conclusion

To find a machine learning method to predict whether a mice PH model is successfully established, a model called HCWOA–KELM is proposed in this paper. To improve the effectiveness and stability of the model, a TRIZ-based hybrid transformation mechanism was added to the WOA. A total of 28 benchmark functions were used to test the global optimality of the improved WOA. The experimental results show that the proposed algorithm performs better global optimization on various test functions than the other eight optimization methods. Similarly, a comparison of bHCWOA–KELM with the other six classifiers and five feature selection methods demonstrates that the proposed model has higher classification accuracy and stability. Finally, the five most selected features in tenfold cross-validation experiments proved crucial for PH prediction.

Acknowledgements This research was supported by grants from the National Natural Science Foundation of China (82003831, 62076185 and U1809209); the Project of Health Commission of Zhejiang Province (2020KY177); the Wenzhou Technology Foundation (Y2020002); the Natural Science Foundation of Zhejiang Province (LZ22F020005), and project funded by the First Affiliated Hospital of Wenzhou Medical University Youth Excellence Project (QNYC114).

Availability of Data and Materials The data involved in this study are all public data, which can be downloaded through public channels.

Declarations

Conflict of Interest The authors declare that there is no conflict of interests regarding the publication of article.

References

1. Mandras, S. A., Mehta, H. S., & Vaidya, A. (2020). Pulmonary hypertension: A brief guide for clinicians. Mayo Clinic Proceedings, 95, 1978–1988.
2. Simonneau, G., Montani, D., Celermaier, D. S., Denton, C. P., Gatzoulis, M. A., Krowka, M., Williams, P. G. & Souza, R. (2019). Haemodynamic definitions and updated clinical classification of pulmonary hypertension. European Respiratory Journal, 53, 01913–2018.
3. Walter, K. (2021). Pulmonary hypertension. JAMA, 326, 1116.
4. Grapsa, J., Pereira Nunes, M. C., Tan, T. C., Cabrita, I. Z., Coulter, T., Smith, B. C., Dawson, D., Gibbs, J. S., & Nihoyannopoulos, P. (2015). Echocardiographic and hemodynamic predictors of survival in precapillary pulmonary hypertension: seven-year follow-up. Circulation: Cardiovascular Imaging, 8, e002107.
5. McLaughlin, V. V., Archer, S. L., Badesch, D. B., Barst, R. J., Farber, H. W., Lindner, J. R., Mathier, M. A., McGoone, M. D., Park, M. H., Rosenson, R. S., Rubin, L. J., Tapson, V. F., & Varga, J. (2009). ACCF/AHA 2009 expert consensus document on pulmonary hypertension. A report of the American College of Cardiology Foundation Task Force on Expert Consensus Documents and the American Heart Association developed in collaboration with the American College of Chest Physicians; American Thoracic Society, Inc.; and the Pulmonary Hypertension Association. Journal of the American College of Cardiology, 53, 1573–1619.
6. Humbert, M., Lau, E. M., Montani, D., Jaïs, X., Sitbon, O., & Simonneau, G. (2014). Advances in therapeutic interventions for patients with pulmonary arterial hypertension. Circulation, 130, 2189–2208.
7. Rubin, L. J. (1993). Primary pulmonary hypertension. Chest, 104, 236–250.
8. Westerhof, B. E., Saouti, N., van der Laarse, W. J., Westerhof, N., & Vonk Noordegraaf, A. (2017). Treatment strategies for the right heart in pulmonary hypertension. Cardiovascular Research, 113, 1465–1473.
9. Chan, L., Chin, L. M. K., Kennedy, M., Woolstenhulme, J. G., Nathan, S. D., Weinstein, A. A., Connors, G., Weir, N. A., Drinkard, B., Lamberti, J., & Keyser, R. E. (2013). Benefits of intensive treadmill exercise training on cardiorespiratory function and quality of life in patients with pulmonary hypertension. Chest, 143, 333–343.
10. Chen, J., Sysol, J. R., Singla, S., Zhao, S., Yamamura, A., Valdez-Jasso, D., Abbassi, T., Shioura, K. M., Sahni, S., Reddy, V., Sridhar, A., Gao, H., Torres, J., Camp, S. M., Tang, H., Ye, S. Q., Comhair, S., Dweik, R., Hassoun, P., ... Machado, R. F. (2017). Nicotinamide phosphoribosyltransferase promotes pulmonary vascular remodeling and is a therapeutic target in pulmonary arterial hypertension. Circulation, 135, 1532–1546.
11. Zelt, J. G. E., Chaudhary, K. R., Cadete, V. J., Mielenzczuk, L. M., & Stewart, D. J. (2019). Medical therapy for Heart failure associated with Pulmonary Hypertension. Circulation Research, 124, 1551–1567.
12. Lahm, T., Douglas, I. S., Archer, S. L., Bogaard, H. J., Chesler, N. C., Haddad, F., Hennes, A. R., Kawut, S. M., Kline, J. A., Kolb, T. M., Mathai, S. C., Mercier, O., Michelakis, E. D., Naeije, R., Tuder, R. M., Ventetulo, C. E., Vieillard-Baron, A., Voelkel, N. F., Vonk-Noordegraaf, A., & Hassoun, P. M. (2018). Assessment of right ventricular function in the research setting: Knowledge gaps and pathways forward. An official American thoracic society research statement. American Journal of Respiratory and Critical Care Medicine, 198, e15–e43.
13. Ayyoubzadeh, S. M., Ayyoubzadeh, S. M., Zahedi, H., Ahmadi, M., & S. R. N. K. (2020). Predicting COVID-19 incidence through analysis of google trends data in Iran: Data mining and deep learning pilot study. JMIR Public Health and Surveillance, 6, e18828.
14. Pourhomayoun, M., & Shahki, M. (2021). Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. Smart Health, 20, 100178.
15. Polat, H., Mehr, H. D., & Cetin, A. (2017). Diagnosis of chronic kidney disease based on support vector machine by feature selection methods. Journal of medical systems, https://doi.org/10.1007/s10916-017-0703-x
16. Abbad Ur Rehman, H., Lin, C.-Y., Mushtaq, Z., & Su, S.-F. (2021). Performance analysis of machine learning algorithms for thyroid disease. Arabian Journal for Science and Engineering, 46, 9437–9449.
17. Pashaei, E., & Pashaei, E. (2022). An efficient binary chimp optimization algorithm for feature selection in biomedical data classification. Neural Computing & Applications, 34, 6427–6451.
18. Alsaeedi, M. A. K., & Kurnaz, S. (2022). Feature selection for diagnose coronavirus (COVID-19) disease by neural network and Caledonian crow learning algorithm. Applied Nanoscience. https://doi.org/10.1007/s13204-021-02159-x
19. Lamba, R., Gulati, T., Alharbi, H. F., & Jain, A. (2022). A hybrid system for Parkinson’s disease diagnosis using machine learning
techniques. *International Journal of Speech Technology*, 25, 583–593.
20. Hu, J., Liu, Y., Heidari, A. A., Bano, Y., Ibrohimov, A., Liang, G., Chen, H., Chen, X., Zaguia, A., & Turabieh, H. (2022). An effective model for predicting serum albumin level in hemodialysis patients. *Computers in Biology and Medicine*, 140, 105054.
21. Faisal, F. U. R., Khatri, U., & Kwon, G.-R. (2021). Diagnosis of Alzheimer’s disease using combined feature selection method. *Journal of Korea Multimedia Society*, 24, 667–675.
22. Hu, J., Han, Z., Heidari, A. A., Shou, Y., Ye, H., Wang, L., Huang, X., Chen, H. L., Chen, Y., & Wu, P. (2022). Detection of COVID-19 severity using blood gas analysis parameters and Harris hawks optimized extreme learning machine. *Computers in Biology and Medicine*, 142, 105166.
23. Heidari, A. A., Mirjalili, S., Faris, H., Aljarahe, I., Mafarja, M., & Chen, H. (2019). Harris Hawks optimization: Algorithm and applications. *Future Generation Computer Systems-the International Journal of Exeience*, 97, 849–872.
24. Yang, Y., Chen, H. L., Heidari, A. A., & Gandomi, A. H. (2021). Hunger games search: Visions, conception, implementation, deep analysis, perspectives, and towards performance shifts. *Expert Systems with Applications*, 177, 114684.
25. Tu, J. Z., Chen, H. L., Wang, M. J., & Gandomi, A. H. (2021). The colony predation algorithm. *Journal of Bionic Engineering*, 18, 674–710.
26. Ahmadianfar, I., Asghar Heidari, A., Gandomi, A. H., Chu, X., & Chen, H. (2021). RUN beyond the metaphor: An efficient optimization algorithm based on Runge Kutta method. *Expert Systems with Applications*, 181, 115079.
27. Ahmadianfar, I., Asghar Heidari, A., Noshadian, S., Chen, H., & Gandomi, A. H. (2022). INFO: An efficient optimization algorithm based on weighted mean of vectors. *Expert Systems with Applications*, 195, 116516.
28. Li, S. M., Chen, H., Wang, M., Heidari, A. A., & Mirjalili, S. (2020). Slime mould algorithm: A new method for stochastic optimization. *Future Generation Computer Systems*, 111, 300–323.
29. Chen, H. L., Yang, B., Wang, J. S., Wang, G., Li, H. Z., & Liu, B. W. (2014). Towards an optimal support vector machine classifier using a parallel particle swarm optimization strategy. *Applied Mathematics and Computation*, 239, 180–197.
30. Han, X., Han, Y., Chen, Q., Li, J., Sang, H., Liu, Y., Pan, Q., & Nojima, Y. (2021). Distributed flow shop scheduling with sequence-dependent setup times using an improved iterated greedy algorithm. *Complex System Modeling and Simulation*, 1, 198–217.
31. Gao, D., Wang, G.-G., & Pedrycz, W. (2020). Solving fuzzy job-shop scheduling problem using DE algorithm improved by a selection mechanism. *IEEE Transactions on Fuzzy Systems*, 28, 3265–3275.
32. Wang, G., Gao, D., & Pedrycz, W. (2022). Solving multi-objective fuzzy job-shop scheduling problem by a hybrid adaptive differential evolution algorithm. *IEEE Transactions on Industrial Informatics*. https://doi.org/10.1109/TII.2022.3165636
33. Li, Q., Chen, H. L., Huang, H., Zhao, X. H., Cai, Z. N., Tong, C. F., Liu, W. B., & Tian, X. (2017). An enhanced grey wolf optimization based feature selection wrapped kernel extreme learning machine for medical diagnosis. *Computational and Mathematical Methods in Medicine*, 2017, 9512741.
34. Cai, Z. N., Gu, J. H., Wen, C. Y., Zhao, D., Huang, C. Y., Huang, H., Tong, C. F., Li, J., & Chen, H. L. (2018). An intelligent Parkinson’s disease diagnostic system based on a chaotic bacterial foraging optimization enhanced fuzzy KNN approach. *Computational and Mathematical Methods in Medicine*, 2018, 2396952.
35. Yu, H., Yuan, K., Li, W., Zhao, N., Chen, W., Huang, C., Chen, H., & Wang, M. (2021). Improved butterfly optimizer-configured extreme learning machine for fault diagnosis. *Complexity*, 2021, 6315010.
36. Ye, X. J., Liu, W., Li, H., Wang, M. J., Chi, C., Liang, G. X., Chen, H. L., & Huang, H. (2021). Modified whale optimization algorithm for solar cell and PV module parameter identification. *Complexity*, 2021, 8878686.
37. Deng, W., Zhang, X. X., Zhou, Y. Q., Liu, Y., Zhou, X. B., Chen, H. L., & Zhao, H. M. (2022). An enhanced fast non-dominated solution sorting genetic algorithm for multi-objective problems. *Information Sciences*, 585, 441–453.
38. Hua, Y. C., Liu, Q. Q., Hao, K. G., & Jin, Y. C. (2021). A survey of evolutionary algorithms for multi-objective optimization problems with irregular pareto fronts. *IEEE/CAA Journal of Automatica Sinica*, 8, 303–318.
39. Zhao, F., Di, S., Cao, J., & Tang, J. (2021). A novel cooperative multi-stage hyper-heuristic for combination optimization problems. *Complex System Modeling and Simulation*, 1, 91–108.
40. Wang, G., Gui, W., Liang, G., Zhao, X., Wang, M., Mafarja, M., Turabieh, H., Xin, J., Chen, H., & Ma, X. (2021). Spiral motion enhanced elite whale optimizer for global tasks. *Complexity*, 2021, 8130378.
41. Chen, C. C., Wang, X. C., Yu, H. L., Zhao, N. N., Wang, M. J., & Chen, H. L. (2020). An enhanced comprehensive learning particle swarm optimizer with the elite-based dominance scheme. *Complexity*, 2020, 4968063.
42. Alshuller, G. S. (1998). 40 Principles: TRIZ Keys to technical innovation.
43. Carrara, P., Morandi, W., Campagiorni, V., & Tandoi, E. (2021). An innovation management framework to enhance TRIZ diffusion in smes. *Acta Technica Napocensis Series- Applied Mathematics Mechanics and Engineering*, 64, 455–464.
44. Petrov, V., & Solodkina, H. (2020). TRIZ: Innovation opportunities for the fashion industry. *Acta Technica Napocensis Series- Applied Mathematics Mechanics and Engineering*, 63, 15–22.
45. Liu, W., Tan, R., Peng, Q., Li, H., Li, Z., & Yang, B. (2020). Impact of TRIZ learning on performance in biologically inspired design. *International Journal of Engineering Education*, 36, 974–987.
46. Hmina, K., El Amine, M., Lasri, L., & Sallaou, M. (2020). Preferences-based approach for TRIZ contradiction matrix exploitation in preliminary design. *Fme Transactions*, 48, 588–599.
47. Spreatico, C. (2021). Quantifying the advantages of TRIZ in sustainability through life cycle assessment. *Journal of Cleaner Production*, 303, 126955.
48. Huang, G. B., Zhou, H., Ding, X., & Zhang, R. (2012). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42, 513–529.
49. Yang, X. S. (2010). A new metaheuristic bat-inspired algorithm. *Springer.
50. Mirjalili, S. (2015). Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowledge-Based Systems*, 89, 228–249.
51. Storn, R., & Price, K. (1997). Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11, 341–359.
52. Jia, D., Zheng, G., & Khurram Khan, M. (2011). An effective memetic differential evolution algorithm based on chaotic local search. *Information Sciences*, 181, 3175–3187.
53. Buchari, M. A., Mardiyanto, S., & Hendradjaya, B. (2018). Implementation of chaotic gaussian particle swarm optimization for optimize learning-to-rank software defect prediction model construction. *Journal of Physics: Conference Series*, 978, 012079.
54. Chen, W., Zhang, J., Lin, Y., Chen, N., Zhan, Z., Chung, H. S., Li, Y., & Shi, Y. (2013). Particle swarm optimization with an aging leader and challengers. IEEE Transactions on Evolutionary Computation, 17, 241–258.

55. Garcia, S., Fernandez, A., Luengo, J., & Herrera, F. (2010). Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. Information Sciences, 180, 2044–2064.

56. Chen, H., Li, S., Heidari, A. A., Wang, P., Li, J., Yang, Y., Wang, M., & Huang, C. (2020). Efficient multi-population outpost fruit fly-driven optimizers: Framework and advances in support vector machines. Expert Systems with Applications, 142, 112999.

57. Hu, J., Liu, Y., Heidari, A. A., Bano, Y., Ibrohimov, A., Liang, G., Chen, H., Chen, X., Zaguaia, A., & Turabieh, H. (2021). An effective model for predicting serum albumin levels in hemodialysis patients. Computers in Biology and Medicine, 2021, 105054.

58. Sankaran, V. G., Xu, J., & Orkin, S. H. (2010). Advances in the understanding of haemoglobin switching. British Journal of Haematology, 149, 181–194.

59. Thom, C. S., Dickson, C. F., Gell, D. A., & Weiss, M. J. (2013). Hemoglobin variants: Biochemical properties and clinical correlates. Cold Spring Harbor Perspectives in Medicine, 3, 11858.

60. Hauser, A., Schmitt, L., Troesch, S., Saugy, J. J., Cejuel-Anta, R., Faiss, R., Robinson, N., Wehrlin, J. P., & Millet, G. P. (2016). Similar hemoglobin mass response in hypobaric and normobaric hypoxia in Athletes. Medicine and Science in Sports and Exercise, 48, 734–741.

61. Deem, S., Kim, J. U., Manjula, B. N., Acharya, A. S., Kerr, M. E., Patel, R. P., Gladwin, M. T., & Swenson, E. R. (2002). Effects of S-nitrosation and cross-linking of hemoglobin on hypoxic pulmonary vasoconstriction in isolated rat lungs. Circulation Research, 91, 626–632.

62. Donadee, C., Raat, N. J., Kianias, T., Tejero, J., Lee, J. S., Kelley, E. E., Zhao, X., Liu, C., Reynolds, H., Azarov, I., Frizzell, S., Meyer, E. M., Donnenberg, A. D., Qu, L., Triulzi, D., Kim-Shapiro, D. B., & Gladwin, M. T. (2011). Nitric oxide scavenging by red blood cell microparticles and cell-free hemoglobin as a mechanism for the red cell storage lesion. Circulation, 124, 465–476.

63. Shaver, C. M., Wickersham, N., McNeil, J. B., Nagata, H., Miller, A., Landstreet, S. R., Kuck, J. L., Diamond, J. M., Lederer, D. J., Kavur, S. M., Palmer, S. M., Wille, K. M., Weinacker, A., Lama, V. N., Crespo, M. M., Orens, J. B., Shah, P. D., Hage, C. A., Cantu, E., 3rd., ..., Ware, L. B. (2018). Cell-free hemoglobin provides primary graft dysfunction through oxidative lung endothelial injury. JCI Insight. https://doi.org/10.1172/jci.insight.98546

64. Shaver, C. M., Upchurch, C. P., Janz, D. R., Grove, B. S., Putz, N. D., Wickersham, N. E., Dikalov, S. I., Ware, L. B., & Bastarache, J. A. (2016). Cell-free hemoglobin: A novel mediator of acute lung injury. American Journal of Physiology: Lung Cellular and Molecular Physiology, 310, L532–L541.

65. Detterich, J. A., Kato, R. M., Rabai, M., Meiselman, H. J., Coates, T. D., & Wood, J. C. (2015). Chronic transfusion therapy improves but does not normalize systemic and pulmonary vasculopathy in sickle cell disease. Blood, 126, 703–710.

66. Fernandez, A. J., & Flaxman, N. A. (1985). Common laboratory tests, values, and interpretations. Special Care in Dentistry, 5, 264–269.

67. Reinhart, W. H. (2016). The optimum hematocrit. Clinical Hemorheology and Microcirculation, 64, 575–585.

68. Crowell, J. W., & Smith, E. E. (1967). Determinant of the optimal hematocrit. Journal of Applied Physiology, 22, 501–504.

69. Ernst, E. (1995). Haematocrit and cardiovascular risk. Journal of Internal Medicine, 237, 527–528.

70. Stauffer, E., Loyrion, E., Hancio, I., Waltz, X., Ulliel-Roche, M., Oberholzer, L., Robach, P., Pichon, A., Brugniaux, J. V., Bouvat, P., Doutreleau, S., Connes, P., & Verges, S. (2020). Blood viscosity and its determinants in the highest city in the world. Journal of Physiology, 598, 4121–4130.

71. Beall, C. M. (2007). Two routes to functional adaptation: Tibetan and Andean high-altitude natives. Proceedings of the National Academy of Sciences of the United States of America, 104, 8655–8660.

72. Bath, P. M., & Butterworth, R. J. (1996). Platelet size: Measurement, physiology and vascular disease. Blood Coagulation & Fibrinolysis, 7, 157–161.

73. Thompson, C. B., & Jakubowski, J. A. (1988). The pathophysiology and clinical relevance of platelet heterogeneity. Blood, 72, 1–8.

74. Gladwin, A. M., & Martin, J. F. (1990). The control of megakaryocyte ploidy and platelet production: Biology and pathology. International Journal of Cell Cloning, 8, 291–298.

75. Kaya, M. G., Yarlioglu, M., Gunebakmaz, O., Gunturk, E., Inanc, T., Dogan, A., Kalay, N., & Topsakal, R. (2010). Platelet activation and inflammatory response in patients with non-dipper hypertension. Atherosclerosis, 209, 278–282.

76. Korniluk, A., Koper-Lenkiewicz, O. M., Kaminska, J., Kemonia, H., & Dymicka-Piekarska, V. (2019). Mean Platelet Volume (MPV): New perspectives for an old marker in the course and prognosis of inflammatory conditions. Mediators of Inflammation, 2019, 9213074.

77. Braeckkan, S. K., Mathiesen, E. B., Njolstad, I., Wilsgaard, T., Stromer, J. H., & Hansen, J. B. (2010). Mean platelet volume is a risk factor for venous thromboembolism: The Tromsø Study, Tromsø, Norway. Journal of Thrombosis and Haemostasis, 8, 157–162.

78. Kovács, S., Csiki, Z., Szőri, K. S., Bereczky, Z., & Shemirani, A. H. (2019). Characteristics of platelet count and size and diagnostic accuracy of mean platelet volume in patients with venous thromboembolism. A systematic review and meta-analysis. Platelets, 30, 139–147.

79. Steiropoulos, P., Papanas, N., Nena, E., Xanthoudaki, M., Goula, T., Froudarakis, M., Pita, E., Maltezos, E., & Bouros, D. (2013). Mean platelet volume and platelet distribution width in patients with chronic obstructive pulmonary disease: The role of comorbidities. Angiology, 64, 535–539.

80. Nadar, S. K., Blann, A. D., Kamath, S., Bevers, D. G., & Lip, G. Y. (2004). Platelet indexes in relation to target organ damage in high-risk hypertensive patients: A study of the Anglo-Scandinavian Cardiac Outcomes Trial (ASCOT). Journal of the American College Cardiology, 44, 415–422.

81. Boos, C. J., Bevers, G. D., & Lip, G. Y. (2007). Assessment of platelet activation indices using the ADVIATM 120 amongst “high-risk” patients with hypertension. Annals of Medicine, 39, 72–78.

82. Inanc, T., Kaya, M. G., Yarlioglu, M., Ardic, I., Ozdogru, I., Dogan, A., Kalay, N., Gunturk, E., Gunebakmaz, O., Gul, I., & Topsakal, R. (2010). The mean platelet volume in patients with non-dipper hypertension compared to dippers and normotensives. Blood Pressure, 19, 81–85.

83. Handtke, S., & Thiele, T. (2020). Large and small platelets—(When) do they differ? Journal of Thrombosis and Haemostasis, 18, 1256–1267.
85. Arslan, D., Cimen, D., Guvenc, O., Gazi, E., Yener, A. U., Temiz, A., Altun, B., Barutcu, A., Erbag, G., & Binnetoglu, E. (2015). Increased platelet distribution width is associated with severity of coronary artery disease in patients with acute coronary syndrome. *Angiology, 66*, 638–643.

86. Vagdatli, E., Gounari, E., Lazaridou, E., Katsibourlia, E., Tsikopoulou, F., & Labrianou, I. (2010). Platelet distribution width: A simple, practical and specific marker of activation of coagulation. *Hippokratia, 14*, 28–32.

87. Detwiler, T. C., Odell, T. T., Jr., & Mac, D. T. (1962). Platelet and Laboratory Medicine, 50, 631–634.

88. Arslan, D., Cimen, D., Guvenc, O., Gazi, E., Yener, A. U., Temiz, A., Altun, B., Barutcu, A., Erbag, G., & Binnetoglu, E. (2015). Increased platelet distribution width is associated with severity of coronary artery disease in patients with acute coronary syndrome. *Angiology, 66*, 638–643.

89. Jindal, S., Gupta, S., Gupta, R., Kakkar, A., Singh, H. V., Gupta, K., & Singh, S. (2011). Platelet indices in diabetes mellitus: Indicators of diabetic microvascular complications. *Hematology, 16*, 86–89.

90. Li, J., & Lin, J. (2020). A probability distribution detection based hybrid ensemble QoS prediction approach. *Information Sciences, 519*, 289–305.

91. Kaito, K., Otsubo, H., Usui, N., Yoshida, M., Tanno, J., Kurihara, K., & Singh, S. (2011). Platelet indices in diabetes mellitus: Indicators of diabetic microvascular complications. *Hematology, 16*, 86–89.

92. Karpatkin, S. (1969). Heterogeneity of human platelets. I. Metabolic and kinetic evidence suggestive of young and old platelets. *Journal of Clinical Investigation, 48*, 1073–1082.

93. Wang, S.-H., Govindaraj, V. V., Górriz, J. M., Zhang, X., & Zhang, Y.-D. (2022). COVID-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network. *Information Fusion, 67*, 208–229.

94. Wang, S.-H., Nayak, D. R., Guttery, D. S., Zhang, X., & Zhang, Y.-D. (2021). COVID-19 classification by CCISNet with deep fusion using transfer learning and discriminant correlation analysis. *Information Fusion, 68*, 131–148.

95. Karpatkin, S., & Strick, N. (1972). Heterogeneity of human platelets. V. Differences in glycolytic and related enzymes with possible relation to platelet age. *Journal of Clinical Investigation, 51*, 1235–1243.

96. Yang, L., Wang, S.-H., & Zhang, Y.-D. (2021). Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network. *Information Fusion, 67*, 208–229.

97. Li, J., Chen, C., Chen, H., & Tong, C. (2017). Towards context-aware social recommendation via individual trust. *Knowledge-Based Systems, 127*, 58–66.

98. Arslan, D., Cimen, D., Guvenc, O., Gazi, E., Yener, A. U., Temiz, A., Altun, B., Barutcu, A., Erbag, G., & Binnetoglu, E. (2015). Increased platelet distribution width is associated with severity of coronary artery disease in patients with acute coronary syndrome. *Angiology, 66*, 638–643.

99. Wang, S.-H., Nayak, D. R., Guttery, D. S., Zhang, X., & Zhang, Y.-D. (2021). COVID-19 classification by CCISNet with deep fusion using transfer learning and discriminant correlation analysis. *Information Fusion, 68*, 131–148.

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