An Intelligent Backup Method for Hospital Information Data Based on Improved Harris Hawk Algorithm

Mengyan Xu
Hospital Information Management Center of the Second Hospital of Anhui Medical University, China
xumengyan0626@163.com

Yanwei Qu
Hefei Haowei Information Technology Co., Ltd
China
18556363510@163.com

Abstract: *Hospital information data exhibits various types and quality issues, including incompleteness and irregularity, complicating backup strategies. Existing methods face challenges such as inadequate multi-objective optimization, limited global search capabilities, and difficulties in balancing efficiency and cost. To address these issues, we propose an intelligent backup method for hospital information data based on an improved Harris Hawk Algorithm (HHA). This method formulates a comprehensive backup strategy fitness function using multi-objective optimization theory, considering key indicators like data integrity, recovery ability, efficiency, security, and cost-effectiveness. The enhanced HHA employs logistic chaotic mapping and elite hierarchy to diversify population initialization, improving global search and population diversity. Additionally, we introduce an adaptive escape energy decreasing strategy and nonlinear jump strength update to enhance exploration and prevent local optima entrapment. Experimental results demonstrate that this method ensures high-quality data backup, significantly boosts backup efficiency, and reduces costs, offering a reliable solution for the secure backup of hospital information data.*

Keywords: *Improved harris hawk algorithm, hospital information data, intelligent backup, multi-objective optimization, logistic chaotic mapping algorithm.*

Received March 24, 2025; accepted October 23, 2025
https://doi.org/10.34028/iajit/23/3/15

1. Introduction

Amidst the swift advancements in information technology and the ongoing refinement of the healthcare system, hospital information data shows an explosive growth trend. These data not only contain patients' diagnosis and treatment records and medical record information, but also involve the hospital's operational data and medical equipment information, which serves as an important basis for the hospital's daily operations and decision-making, possessing high importance and sensitivity [5]. These data are directly related to patients' life safety, hospital operations and management, and medical quality. Once the data are lost or damaged, it will bring immeasurable losses to both the hospital and patients. Therefore, intelligent backup of hospital information data is an important measure to protect hospitals' normal operations and patient safety. Data backup is essential to ensure the security and reliability of hospital information. However, hospital information data has the characteristics of diversity, complexity and high sensitivity, which makes data backup work face enormous challenges. Traditional data backup methods often employ a single backup strategy and simple backup tools, making them difficult to adapt to the complexity and diversity of hospital information data [24]. Additionally, traditional backup methods usually require manual intervention, which is cumbersome and error-prone.

In recent times, the remarkable progression of artificial intelligence technology has been evident [14], intelligent backup methods have gradually become a research hot spot. These methods use machine learning, optimization algorithms and other technical means to achieve intelligent backup and management of hospital information data. Through the introduction of intelligent technology, automated backup, intelligent management and efficient recovery of hospital information data can be realized to improve backup efficiency and data security. However, although these methods have achieved certain research results, there are still some problems, such as unstable performance of algorithms and poor optimization effect, which limit their wide promotion in practical applications.

Researchers have delved into these challenges and proposed corresponding solutions, so that machine learning technology and optimization algorithms can be more widely used in data backup, medical decision support, patient health monitoring and other key scenarios [15]. To promote the development of related technologies and provide ideas and methods for the intelligent management of hospital information data. In [17] a wrapper based Binary Improved Grey Wolf Optimizer (BIGWO) algorithm was proposed to address the various complex symptoms, significant individual differences, and unclear pathogenesis of Parkinson's disease, which make it difficult to select relevant features of the disease. This algorithm encodes the feature search space using five different transfer functions and refines the search process through mutation operations within BIGWO. This method determines the optimal features for disease diagnosis and ensures their subsequent backup. This algorithm is prone to getting stuck in local optima when dealing with high-dimensional medical data, especially when dealing with biomedical data with nonlinear and high noise characteristics such as Parkinson's disease. Its ability to maintain population diversity is insufficient, resulting in a significant decrease in the efficiency of feature subset search. In [21] a hybrid ensemble classifier based on pre-trained neural network feature extraction is proposed for accurately identifying the classification and degree of cervical cell dysplasia, and backing up based on the final classification results. However, when faced with dynamically changing multi-source heterogeneous medical data in hospital information data backup scenarios, static pre trained features are difficult to adaptively adjust, resulting in insufficient ability to recognize newly emerging data patterns. In [18] a multi-objective selection method based on Particle Swarm Optimization (PSO) algorithm is proposed to solve the problems of reduced disease diagnosis accuracy and increased computational complexity caused by high-dimensional features in medical datasets. This method displays the original features in the form of a graphical representation model, calculates the feature concentration of all nodes in the graph, and uses an improved PSO particle swarm optimization algorithm for final feature selection to complete the corresponding data backup. There are a large number of implicit complex feature interactions in medical data, and traditional PSO algorithms based on euclidean distance particle movement mechanism are difficult to effectively model such high-order correlations. When handling hospital information data backup tasks, the linear concentration calculation method may lose the deep semantic associations between key features, resulting in the final selected feature subset not fully reflecting the true information structure of medical data, which affects the integrity and availability of backup data. In [1] a new variant of Grey Wolf Optimizer (GWO) is proposed to address potential issues with the GWO and most of its variants. This method combines memory, asymptotic operators, and random local search techniques to accelerate convergence and obtain the optimal solution. When dealing with large-scale tasks such as hospital information data backup that require strict timeliness and integrity, its random search mechanism is difficult to effectively balance the relationship between global exploration and local development, resulting in premature convergence or computational redundancy of the algorithm when facing high-dimensional heterogeneous medical data, which affects the optimality and execution efficiency of the backup plan.

The Harris Hawks Optimization (HHO) methodology represents a novel bio-inspired optimization technique, mirroring the hunting tactics employed by Harris Hawks in nature [22]. Its core idea is to achieve effective solution to complex multidimensional problems by simulating the unique predatory characteristics and foraging strategies of Harris Hawks. Throughout the algorithm's execution, the Harris Hawk, acting as a potential solution, progressively converges towards the prey's location (analogous to the optimal solution of the problem), thereby discovering the most effective solution iteratively.

The existing methods commonly suffer from insufficient capture of the dynamic characteristics and complex associations of medical data backup. A hospital information data intelligent backup method based on improved Harris Hawks algorithm is proposed to address this common deficiency. By introducing adaptive search mechanisms and intelligent optimization strategies, deep mining of multidimensional features of medical data has been achieved. The key technological breakthrough lies in the establishment of a dynamic response model that can perceive real-time changes in data patterns and optimize the feature selection process using hybrid intelligent algorithms. This method effectively solves the problems of slow convergence speed and susceptibility to local optima that traditional algorithms encounter when dealing with high-dimensional heterogeneous medical data, significantly improving the integrity and execution efficiency of backup solutions.

2. Target Determination of Hospital Data Information Backup Strategy Based on MOP Model

For the purpose of solving the problem of insufficient multi-objective optimization and ensuring the reliability of the backup strategy for the final hospital data information, a comprehensive and efficient backup strategy fitness function is constructed through Multi-Objective Optimization Theory (MOP), taking into account the integrity, recovery ability, backup efficiency, security, cost-effectiveness, and other nonlinear key indicators of the hospital data information, providing support for subsequent solutions.

2.1. MOP Planning Equation

MOP is used to deal with problems involving two or more objective functions being optimized simultaneously [3, 16]. In MOP, some objective functions complement or contradict each other, and the selection of decision variables affects multiple objective function values. Therefore, practical application requires holistic evaluation and integration of various objective functions, finding a balance between them while maximizing the system's overall objective function value. The mathematical formulation is as follows:

$$f(x)=[f_1(x),f_2(x),\cdots f_n(x)] \quad (1)$$

$$s.t.\begin{cases} g_i(x)\le 0 & i=1,2,\cdots m \\ h_j(x)=0 & j=1,2,\cdots n \end{cases} \quad (2)$$

Among them, $f(x)$ is the objective function; $f_i(x)$ serves as a subsidiary objective function; $g_i(x)$ is an inequality constraint; $h_j(x)$ is the equation constraint.

2.2. Evaluation Metrics for Nonlinear Multimetric Backup Strategies

Hospital data information is rich and diverse, covering medical imaging data [7], vital signs data [19], health management data, health resources data, public health data [10], case data [12], laboratory test data [9], medical payment data and other aspects. The backup strategy for hospital data and information is a key measure to ensure the stable operation of hospital information systems and data security. To evaluate the effectiveness of a hospital data and information backup strategy, it is necessary to comprehensively consider multiple indicators such as data integrity, recovery capability, backup efficiency, security, and cost-effectiveness [2], for the purpose of ensuring that hospital data are safe, reliable, and effectively utilized.

2.2.1 Data Integrity Index (DI)

Backup data is complete, no loss or damage, which is the most basic indicators to evaluate the quality of backup. Its mathematical model can be abbreviated as:

$$DI=\frac{\sum_{i=1}^{n}D_i}{\sum_{i=1}^{n}T_i} \quad (3)$$

Among them, D_i indicates the amount of data that successfully backed up for the *i* th time. T_i indicates the total amount of data that should be backed up for the *i th* time. *n* indicates the number of backups.

2.2.2 Data Recovery Capability Index (DRC)

Recoverability, as a core metric of backup strategy, involves complete and partial recovery of data, as well as data consistency and accuracy during the recovery process. Its mathematical model can be abbreviated as follows:

$$DRC=\frac{\sum_{j=1}^{m}R_j}{\sum_{j=1}^{m}A_j} \quad (4)$$

Among them, R_j indicates the amount of data successfully recovered for the *j* th time, A_j indicates the amount of data recovered for the *j th* attempt, m indicates the number of recoveries.

2.2.3 Backup Efficiency Index (BE)

Backup speed is critical to the efficiency of hospital data recovery. In order to maintain normal operation, the backup strategy needs to minimize the backup time and system resource usage. The specific representation is as follows:

$$BE=\frac{\rho\times\Delta t}{\sum_{k=1}^{\rho}T_k}\otimes\kappa \quad (5)$$

Among them, T_k indicates the time required for the *k th* backup, ρ indicates the number of backups, Δt indicates the scheduled backup cycle time. *K* indicates the percentage of system resources.

2.2.4 Security Index (S)

Security is an integral part of hospital data and information backup. The backup data itself needs to be properly protected against unauthorized access, tampering, or deletion. Security is a comprehensive indicator that can be evaluated through the following sub-indicators, with corresponding weights and scores assigned.

- Encryption Rate (ER): the ratio of the amount of data backed up using encryption algorithms to the total amount of data backed up, with a value range of 0-1.
- Access Control (AC): restrict access to backup data through authentication and permission management. The default value is determined by the scoring system.
- Storage Security (SS): the inability to store backup data or the security assessment of the cloud environment [11]. The default value is determined according to the scoring system.

The integrated assessment formula can be expressed as follows:

$$S=\omega_1\times ER+\omega_2\times AC+\omega_3\times SS \quad (6)$$

Among them, ω_1, ω_2, ω_3 are the weights of each sub-indicator, respectively, and $\omega_1+\omega_2+\omega_3=1$.

2.2.5 Cost Effectiveness Index (CE)

Storage cost is an important factor for hospitals to consider when implementing a backup strategy. An efficient backup strategy should ensure data security and integrity while minimizing storage costs. This includes investments in hardware equipment, maintenance costs, and storage space utilization. A quantitative assessment of this can be summarized as follows:

$$CE=\frac{C_{savings}}{C_{total}} \quad (7)$$

Among them, $C_{savings}$ indicates the cost savings or losses avoided through the backup strategy, C_{total} indicates the total cost of the backup strategy.

2.3. Objective Function Construction for Data Backup Strategy

Due to the complexity and sensitivity of hospital data, the backup strategy needs to flexibly adapt to different data characteristics and business needs. In this paper, based on the five indicators of data integrity, recovery capability, backup efficiency, security and cost-effectiveness, as a sub-objective function in the above equation (1), to build the backup strategy objective function, as shown in the following equation (8). The relationship between the indicators is complex, and should be emphasized according to the actual situation and needs. When data changes frequently, data integrity and recovery capabilities should be emphasized; high real-time requirements, the need to improve backup efficiency; facing network security threats, the need to ensure data security. At the same time, cost-effectiveness should not be ignored. Therefore, we need to find the optimal backup strategy to ensure the best indicators. In practice, we need to adjust the strategy according to the data changes and business needs, based on the 2.1 theory, by assigning weights to each backup strategy evaluation index, transform the original multidimensional optimization challenge into a single objective optimization problem. Subsequently, an efficient single objective optimization algorithm was adopted to address this transformed problem, aiming to build an unbreakable data security defense line for the stable operation of the hospital system, ensuring that business continuity is not affected.

$$\begin{aligned} f(x) &= \sum_{i,j,k=1}^{n,m,\sigma}(\phi_1 \times DI(x) + \phi_2 \times DRC(x) \\ &+\phi_3 \times BE(x) + \phi_4 \times S(x) + \phi_5 \times CE(x)) \end{aligned} \tag{8}$$

Among them, ϕ_1, ϕ_2, ϕ_3, ϕ_4, ϕ_5 are the weight values of each assessment sub-indicator, respectively, which are preset and adjusted according to the actual situation, and $\phi_1 + \phi_2 + \phi_3 + \phi_4 + \phi_5 = 1$.

3. Solving the Objective Function of a Data Backup Policy Based on the Harris Hawk Algorithm

Based on the objective function of the data backup strategy built above, to obtain the optimal intelligent backup scheme for hospital information data, the HHO algorithm is employed for solution. The Harris Hawk Optimization methodology represents a robust and adaptable approach to solving optimization problems. The main process of the algorithm can be divided into three stages: global exploration, transition, and local exploitation [6]. In the exploration phase, Harris Hawks fly randomly in the search space to find potential prey (hospital information data backup schemes). In the transition phase, Harris Hawks adjust their flight strategy according to the escape energy and behavior patterns of prey for more effective hunting [13, 23]. In the exploitation stage, Harris Hawks employ various predation strategies, such as siege and surprise attacks, to accurately locate and capture prey. The specific description of the solution process for the hospital information data backup scheme at each stage of the Harris Hawk algorithm is as follows:

3.1. Global Exploration Phase

The Harris Hawk tracks and detects prey with its powerful vision and stops at random locations to wait and detect prey. When the prey escapes energy $|E|\geq 1$, the algorithm performs the global search behavior, its expression is as follows:

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1 |X_{rand}(t) - 2r_2 X(t)| & q \geq 0.5 \\ [X_{rabbit}(t) - X_m(t)] - r_3[lb + r_4(ub - lb)] & q < 0.5 \end{cases} \tag{9}$$

Among them, $X(t+1)$ and $X(t)$ are the position of the eagle at the $t+1$ and t iteration respectively; the $X_{rand}(t)$ is the randomized position of the eagle at the t iteration; the $X_{rabbit}(t)$ is the current optimal individual position; the $X_m(t)$ denotes the location of the center of the population. γ and q are random numbers between [0, 1]. *ub* and *lb* indicate the upper and lower limits of the search space.

3.2. Conversion Phase

The transition phase dynamically alternates between distinct exploitation strategies, contingent upon the evasion energy exhibited by the targeted prey. During the iterative search, the behavior of the algorithm in performing a global exploration or local development depends on the magnitude of the E-value of the linear decreasing prey energy, which is expressed as:

$$E = 2E_0(1 - \frac{t}{T}) \tag{10}$$

Among them, T is the maximum number of iterations; $E_0 \in [0, 1]$ is the initial energy value of the prey.

3.3. Localized Development Phase

When the prey escapes energy $|E|<1$, the algorithm performs local exploitation behavior. The Harris Hawk form four roundup strategies based on the random number in the roundup r and prey energy $|E|$ compare with the size of 0.5.

3.3.1. Soft Seining

When $|E|\geq 0.5$ and $r\geq 0.5$, the prey has plenty of energy E to escape, the eagles implement a soft roundup strategy, and the eagle's position is updated by the formula:

$$X(t+1) = \Delta X(t) - E\,|JX_{rabbit}(t) - X(t)| \tag{11}$$

Among them, J represents prey jumping energy, take a random number between [0, 2].

3.3.2. Hard Seining

When $|E|<0.5$ and $r\geq 0.5$, the escape energy E of the prey is lower, without sufficient energy to escape, the eagles execute a hard roundup strategy, and the eagle's position is updated with the formula:

$$X(t+1) = X_{rabbit}(t) - E\,|\Delta X(t)| \tag{12}$$

3.3.3. Rapid Swooping Soft Seining

When $|E|\geq 0.5$ and $r<0.5$, the hawks adjust their positions utilizing Equation (13), subsequently contrasting these new positions with the current fitness function values to assess improvement, if the fitness value is not improved, it means that the roundup fails, then the hawks randomized wandering based on *Levy*, using Equation (14) for position updating.

$$Y = X_{rabbit}(t) - E\,|JX_{rabbit}(t) - X(t)| \tag{13}$$

$$Z = Y + S \times Levy(D) \tag{14}$$

Among them, S is z random variables with the dimension of D, Levy is the flight function.

3.3.4. Rapid Swooping Hard Roundup

When $|E|\geq 0.5$ and $r<0.5$, the prey lacks sufficient evasive capabilities, the eagle group executes the fast dive hard roundup strategy, and its position is updated as shown in Equation (15). If the fast dive fails, then execute *Levy* randomized wanderings and position updates using Equation (15).

$$Y = X_{rabbit}(t) - E\,|JX_{rabbit}(t) - X_m(t)| \tag{15}$$

4. HHO Algorithm Combination Improvement Strategy

Based on the above process, to enhance the algorithm's global search capability and initial population diversity, enabling better adaptation to hospital information data backup problem characteristics, the Harris Hawk algorithm is improved through the introduction of logistic chaotic mapping and elite hierarchy. Incorporating an escape energy decreasing update strategy and nonlinear jump strength update strategy allows the algorithm to adaptively adjust search ranges and strengthen exploration capability, thereby avoiding local optima and obtaining the optimal intelligent backup scheme for hospital information data.

4.1. Initial Stock Diversification Improvement

4.1.1. Logistic Chaotic Mapping

Conventionally, the initial data set for the HHO algorithm is typically constructed through a process of random generation, which makes the diversity of the population extremely uncertain, and the spatial distribution may be extremely uneven, resulting in poor optimization effect of the algorithm. Therefore, the logistic chaotic mapping method is introduced to improve the HHO algorithm for intelligent backup of hospital information data. Logistic chaotic maps can yield initial populations that exhibit both randomness and ergodicity, fostering diversity within the algorithm and, subsequently, augmenting its capability for global exploration. This helps the algorithm to cover hospital information data more comprehensively, reduce the risk of data loss, and better handle incomplete and inaccurate data. Set the objective function as Equation (16), and solve the cascaded chaotic sequence $[y_n]$ through Equation (17):

$$\min f(x_1, x_2, \cdots, x_n), lb_i < x_i < ub_i \tag{16}$$

$$\begin{cases} x'_{n+1} = 4x'_n(1-x'_n), \\ y'_n = \dfrac{1}{\pi} arcsin(2x'_{n+1}-1) - \dfrac{1}{2}, \\ x_{n+1} = 4y'_n(1-y'_n), \\ y_n = \dfrac{1}{\pi} arcsin(2x_{n+1}-1) - \dfrac{1}{2}n \end{cases} \tag{17}$$

Among them, n is the population size. lb_i and ub_i denote the upper and lower bounds of the solution space. The linear transformation is performed to $[y_n]$, as shown in Equation (18), to obtain the initial position of the Harris's hawk:

$$P_i = lb_i + (ub_i - lb_i)y_n \tag{18}$$

4.1.2. Elite Hierarchy

As iterations progress, population diversity tends to diminish, predisposing the algorithm to premature convergence toward local optima, thereby compromising convergence accuracy and potentially yielding suboptimal solutions. To enhance global search capability and mitigate diversity reduction in later iterations, an elite hierarchy is introduced while applying logistic chaotic mapping technology. The elite hierarchy gradually improves population diversity through continuous iteration and optimization, enabling more comprehensive coverage of hospital data information during the search process, particularly sensitive and confidential information. By preserving and inheriting characteristics of outstanding individuals, the elite hierarchy facilitates faster discovery of global optima. Consider enhancing suboptimal solution information exchange during iterations by strategically selecting three elite positions to replace current optima, thereby guiding other individuals toward more promising search directions. This can be expressed as:

$$X_{rabbit} = \sum_{j=\alpha,\beta,\gamma} \frac{f(X_{jbest}(t))}{\sum_{z=\alpha,\beta,\gamma} f(X_{zbest}(t))} X(t) \tag{19}$$

Among them, X_{jbest} is the current dominant individual in the population, $f(X_{zbest})$ is the fitness value of the dominant individual of the current population.

4.2. Escape Energy Decreasing Update Strategy

In the traditional HHO algorithm, the escape energy factor E controlling the transition of the algorithm from global to local, the larger value of E indicate *a* larger search scope, i.e., global search, while smaller values indicate a smaller search scope, i.e., local search. But the updating method of E is linearly decreasing from 2 to 1, i.e., its value is small in the latter half, the search range is small, and it is easy to fall into the local extreme value. Given the aforementioned issues, this paper proposes a nonlinear energy factor *E*, avoiding premature convergence in the late iterations and increasing the search accuracy of the algorithm, the improved energy factor E can be expressed as:

$$E = 2E_0\omega(1-t/T) \tag{20}$$

$$\omega = \varpi[\lambda - (\lambda-\gamma)\times\frac{1}{e}\times(e^{t/T}-1)] \tag{21}$$

Among them, λ is the initial value of the weights, γ is the final value of the weights, t is the current iteration number, ϖ is a random number between [0, 1], T is the maximum number of iterations.

4.3. Nonlinear Jump-Strength Updating Strategy

In the traditional HHO algorithm, the random numbers of the jumping strength of prey $J\in[0, 2]$ results in the algorithms are very difficult to have a good convergence in the optimization process, which leads to the reduction of the optimization accuracy. Considering that the jumping strength is greatly affected by the energy decay in the late iteration, the jumping strategy of Equation (22) is proposed:

$$J = 2r(e^{1-\sqrt{\frac{1+t}{T}}}-1) \tag{22}$$

Among them, J is the jumping strength of the prey, T is the maximum number of iterations, $r\in[0, 1]$, t is the current iteration number, $t=0,1,2,L,T-1$.

In a real hospital IT system environment, the improved HHO algorithm dynamically perceives the load status and storage resource distribution of backup servers, and optimizes data partitioning strategies in real time. When the Picture Archiving And Communication System (PACS) system generates new medical images, the algorithm first evaluates the I/O throughput, remaining capacity, and network latency of each storage node, and then adaptively adjusts the distribution weights of data blocks based on the non-linear jump strength update strategy. For emergency medical record data, the algorithm automatically increases its backup priority and uses multi copy redundant encoding to ensure high availability. Continuously monitor the health status of the storage cluster during the backup process. If a disk failure or network anomaly is detected, immediately trigger the data migration mechanism to redistribute the affected data blocks to the optimal node. Initiate global consistency verification during low load periods in the early morning of each day, eliminate silent errors through verification and comparison, and ensure the long-term integrity of clinical data.

5. Intelligent Backup of Hospital Data Information Based on Improved HHO

The hospital Information Technology (IT) system needs to build a hybrid cloud architecture as the foundation for HHO algorithm deployment. The core requirements include: real-time data perception layer to collect and store node status, distributed computing cluster to run optimization algorithms, and automated strategy engine to execute dynamic backup instructions. The deployment architecture is divided into three layers: the edge device layer processes real-time data blocks such as PACS, the private cloud layer runs an improved HHO algorithm for multi-objective decision-making, and the hybrid storage layer allocates hot and cold data to Solid-State Drive (SSD)/mechanical disk/encrypted storage according to weight. The main challenge lies in the quantification bias caused by the heterogeneity of medical data, which needs to be addressed through data standardization preprocessing; There is a contradiction between the real-time requirements of the algorithm and the delay of the hospital network, and edge computing nodes need to be deployed to compress the decision delay; Security compliance requires embedding a zero trust architecture to ensure that encryption rates and access control policies comply with standards such as Health Insurance Portability And Accountability Act (HIPAA).

As medical data volumes proliferate and information technology progresses rapidly, ensuring the safeguarding and backup of hospital data and information becomes paramount in significance [8, 20]. Traditional backup methods are often characterized by inefficiency, high cost, and difficulty in coping with complex and changing data environments. Therefore, this paper proposes an intelligent backup method for hospital data information based on the improved Harris hawk algorithm and backup strategy.

In terms of backup strategy, the integrity of hospital data and information, recovery capability, backup efficiency, security and cost-effectiveness are taken into account. Through the development of scientific backup plans and strategies to ensure that hospital data can be quickly recovered in the event of an accident, while reducing the cost and time expenses in the backup process. In addition, a dynamic adjustment mechanism has been introduced to adjust the backup strategy in real time according to the changes in hospital data and backup requirements to ensure the timeliness and effectiveness of the backup work.

Based on the core idea of improved HHO algorithm:

- Global exploration stage: during the initial phase of data backup, the refined HHO algorithm is employed to assess and categorize the data types, quantity, importance, sensitivity and other relevant characteristics of the hospital's internal data. According to the evaluation results, different backup strategies can be developed for different types of data, such as full backup, incremental backup, real-time backup, etc.

- Conversion phase: according to the change of data, the health of the storage device, network bandwidth and other factors, adjust the backup frequency and backup mode at the right time, and utilize an adaptive weighting mechanism to dynamically modulate the weight factor within the backup strategy's objective function.
- Local development stage: determine the recovery precedence based on the significance of data and the Recovery Time Objective (RTO), ensuring prompt restoration of vital data in the event of loss or corruption. Simultaneously, multiple recovery strategies can be introduced, such as remote recovery, local recovery, snapshot recovery, etc., to meet different recovery scenarios and requirements.

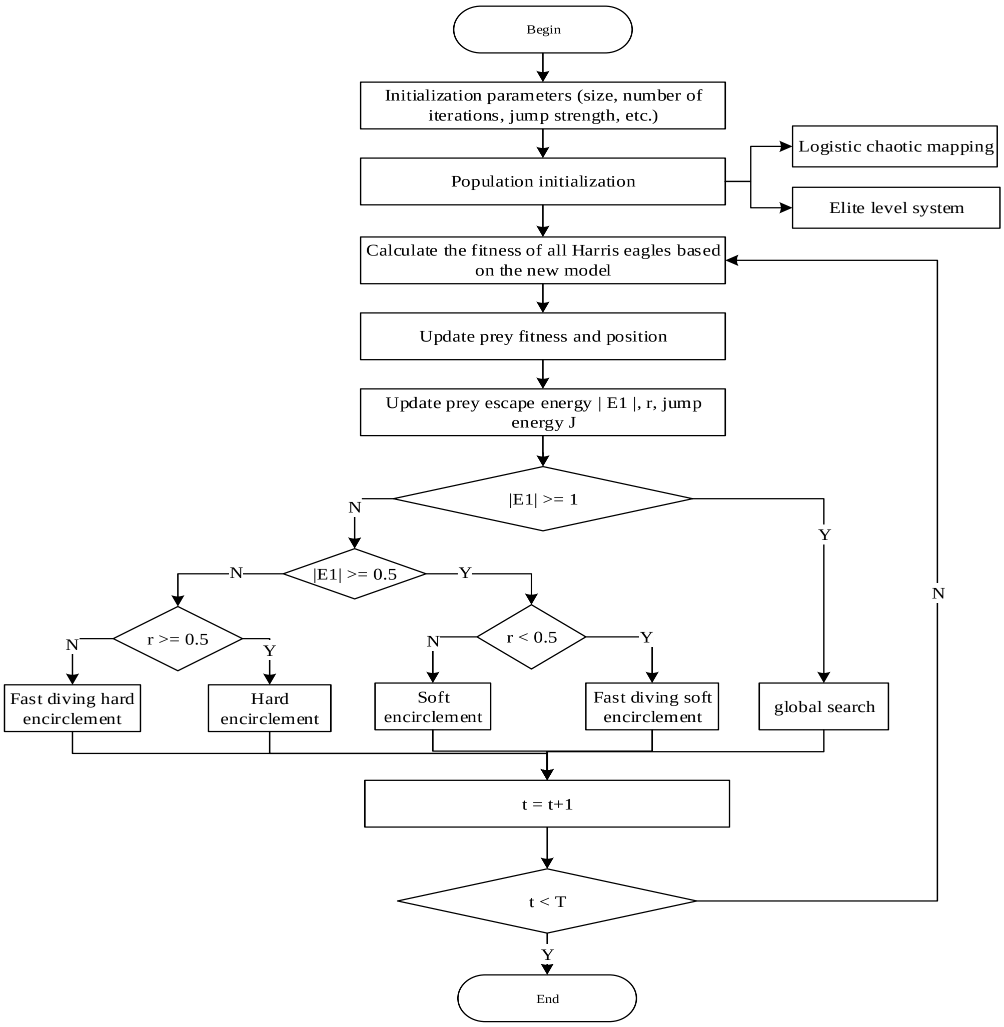


Figure 1. Framework of HHO.

The core of this method is to combine the multi strategy improved HHO algorithm with the backup strategy to realize the intelligent backup of hospital data information. The specific steps of the algorithm are:

- *Step* 1: Input hospital data information, including case data, medical image data, laboratory test data, vital signs data, health management data, medical payment data, health resources data, public health data and other aspects, and determine the set initialization parameters, including population size, maximum number of iterations, switching probability, jumping intensity, etc.
- *Step* 2: Initialize the population using logistic chaotic mapping algorithm and elite hierarchy.
- *Step* 3: Based on the MOP theory, combined with the five indicators of integrity, resilience, backup efficiency, security and cost-effectiveness, build an objective function for the intelligent backup strategy of hospital data information, replace it with the fitness function in the traditional HHO algorithm, calculate the objective function value of each Harris Hawk location according to this function, and determine the individual fitness value, Get the current optimal function value and corresponding position.
- *Step* 4: Revise the escape energy and prey's jumping intensity, leveraging the computation of a nonlinear escape energy factor to facilitate a seamless transition between global exploration and localized search, thereby enhancing the algorithm's intelligence.

- *Step* 5: Select the updated strategy contingent upon the prey's escape energy. If the absolute escape energy exceeds 1, the eagle group embarks on a search phase. Conversely, when the absolute escape energy is less than 1, the eagles transition to a capture phase. The threshold of 0.5 in the absolute escape energy dictates whether the eagle group resorts to a circling capture or a sudden, intense attack. In the event that the prey evades capture, the eagle group initiates a gradual, progressive pursuit strategy.
- *Step* 6: Evaluate and compare the fitness scores of all individuals across the entire search domain, constantly update the current global optimal solution, and judge whether the maximum iteration number T is reached. If the conditions are met, go to step 7, otherwise go to step 2.
- *Step* 7: Output optimal fitness and position X_{rabbit}, this optimal solution represents the best backup strategy under current conditions. The hospital can formulate a specific backup plan based on this optimal solution, including determining the backup time point, selecting the appropriate backup method (such as full backup, incremental backup or differential backup), and determining the storage location of backup data (mechanical disk, SSD disk, encryption disk, ordinary disk, etc.).

In order to show this process more intuitively, based on the information of the above steps, the flowchart of intelligent backup of hospital data and information is drawn in a concise and clear graphical way [4], as shown in Figure 1 above.

6. Experimental Analysis

6.1. Experimental Preparation

6.1.1. Hardware and Software Platforms

For the purpose of verifying the effectiveness and robustness of the improved Harris Hawk algorithm of multi strategy fusion and the intelligent backup of hospital data information based on the algorithm and backup strategy, this experiment uses high-performance computer configuration. The experimental equipment is equipped with Intel Core i7 multi-core processors to ensure efficient implementation of algorithms; 16GB Double Data Rate (DDR4) RAM, meeting the memory requirements for processing large amounts of data; 512GB solid state hard disk ensures data read/write speed and storage capacity. The stable and reliable Windows 11 is selected as the operating system, and the simulation platform Matlab 2021a and related data processing libraries are installed. In addition, the experimental environment also includes a stable network connection to ensure the timeliness and accuracy of data backup and transmission. This configuration provides a strong guarantee for the smooth progress of the experiment. The experimental environment is shown in Figure 2 below.

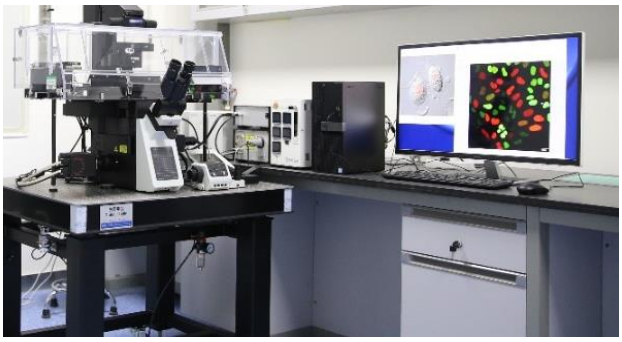


Figure 2. Schematic diagram of the experimental environment.

6.1.2. Comparison Algorithm and Initial Settings

For the purpose of verifying the performance superiority of the multi strategy improved HHO algorithm in this paper, GWO, PSO and traditional HHO are introduced for comparison. See Table 1 for details of the initial settings of the comparison algorithm:

Table 1. Comparison algorithm and initialization settings.

Algorithm	Convergence factor
GWO	Parameter reduced from 3 to 0
PSO	C=0.5, w=0.8
GA	P1=1, P2=0.3
HHO	N=30, T=300, D=30

6.1.3. Standardized Test Functions

For the purpose of verifying the superiority of this multi strategy improved HHO algorithm in performance, some international standard test functions are used for comparative experiments. F1~F3 are unimodal test functions, which are applicable to the optimization speed and local optimization accuracy of the detection algorithm. F4~F6 are multimodal test functions, and F7~F9 are fixed dimension multimodal test functions, which are suitable for assessing the global optimization capabilities of the detection algorithm. The designated test set for this function is detailed in Table 2.

Table 2. International standard test function.

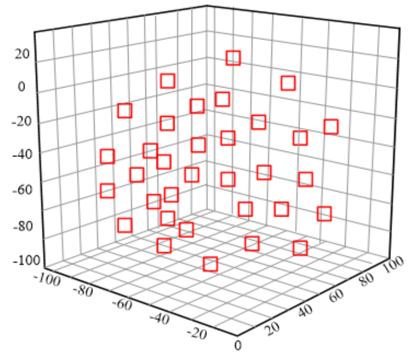
Function	Name	Value Range	Dimension	Optimal solution
F1	Sphere	[-100,100]	N	0
F2	Ellipsoidal	[-100,100]	N	0
F3	Quartic	[-100,100]	N	0
F4	Rastrigin	[-5.12,5.12]	N	0
F5	Ackley	[-32,32]	N	0
F6	Griewank	[-600,600]	N	0
F7	Foxholes	[-65,65]	2	3.08E-04
F8	Six-Hump	[-5,5]	2	-1.031628
F9	Branin	[-5,5]	2	0.398

6.2. Experimental Design and Results

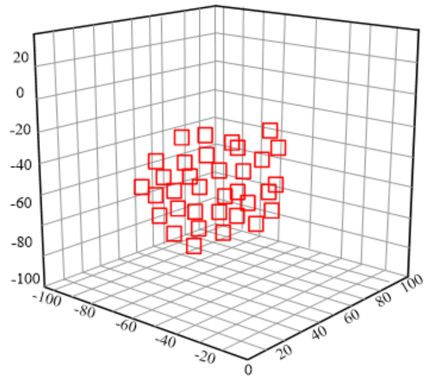
6.2.1. Analysis of the Effectiveness of the Proposed Algorithm

1) Comparative assessment of improvement strategies for population initialization

For the purpose of verifying the effectiveness of population initialization based on logistic chaotic mapping algorithm and elite hierarchy system in this paper, population initialization and related parameter settings are carried out at the same time under the premise of the same search space and number of samples, and the random initialization population in traditional HHO algorithm is compared with the population initialization of this algorithm, as shown in Figure 3.



a) Original initializing population.



b) Logistic chaos initialization population.

Figure 3. Initialize population.

From Figure 3, it is easy to see that, first of all, in terms of population diversity, the population initialized by Logistic chaotic map shows higher uniformity and dispersion. Due to the ergodicity and randomness of Logistic chaotic map, it can generate more uniform initial population in the search space. In contrast, the problem of uneven distribution often exists in the randomly initialized population, which may lead the algorithm to prematurely converge to a local optimum during the initial search phase. Secondly, in terms of convergence speed, the initialization population of logistic chaotic map also shows a faster convergence speed. Because the elite system mechanism is introduced into the algorithm in this paper, the quality of the initial population is higher, and the algorithm can approach the global optimal solution faster in the search process. However, due to the uneven quality of initial solutions, the random initialization population often needs more iterations to achieve the same convergence effect. Evidently, by integrating logistic chaotic mapping and an elite system during population initialization, the proposed algorithm significantly enhances population diversity and accelerates the convergence rate.

2) Comparative evaluation of nonlinear improvement strategies for fugitive energy factor.

For the purpose of improving the traditional algorithm which tends to fall into localized search in the late iteration, this paper proposes the nonlinear improvement strategy of an escape energy factor El According to Equation. (18), it can be concluded the updated manner of El, i.e., $El=E\times E_0$, of which, E_0 denotes the initial energy of the prey. Through the experimental parameter setting, the change curve of the energy factor E and the escape energy factor El can be obtained at the iteration times T=100, as shown in Figures 4 and 5.

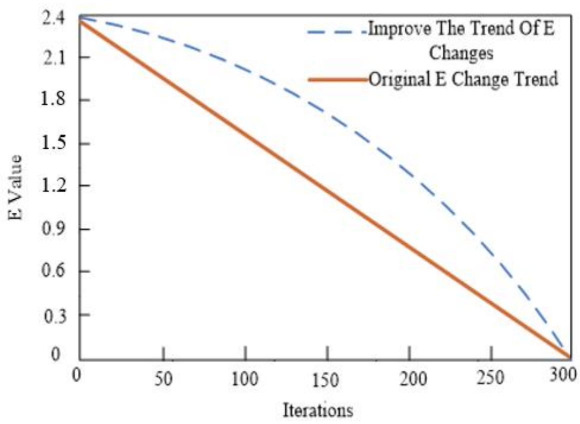
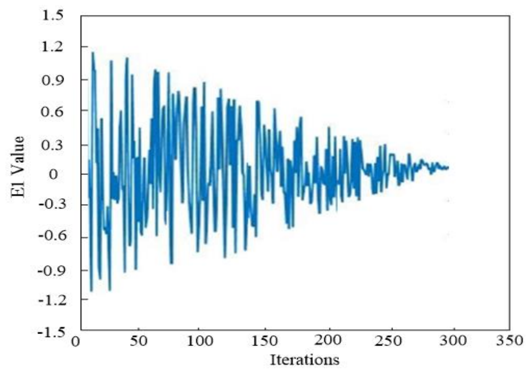
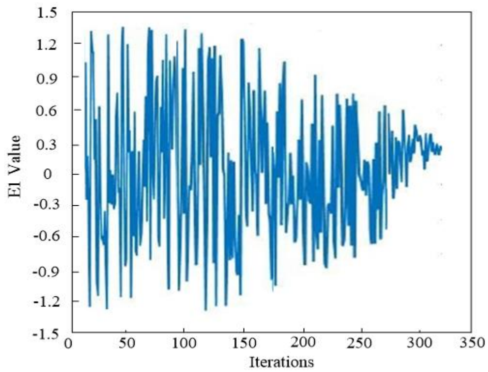


Figure 4. E variation curve graph.



a) Changes in the original cscape energy factor E1.



b) Improved escape energy factor E1 variation chart.

Figure 5. Escape energy factor El variation curve.

Figure 4 illustrates that the refined E parameter exhibits a gradual decline in the algorithm's initial stages, fostering a wider global search scope and mitigating “premature convergence”. Subsequently, the rate of E decrease accelerates, bolstering local search capabilities and enhancing overall search efficiency. This approach ensures that the HHO algorithm primarily focuses on global exploration early on, transitioning to a balanced combination of global and local searches in its intermediate stages. In the later stage, the algorithm retains the possibility of global search on the premise of ensuring local search.

Through the comparison of Figure 5, it is definitively discerned that the refined escape energy factor El undergoes a gradual variation in the algorithm’s early stages, fostering improved global search capabilities during initialization. Conversely, in later stages, El diminishes more rapidly, enhancing local search proficiency and search efficiency. As a result, the methodology presented herein adeptly balances global and local search strategies, enabling the HHO algorithm to execute both global and local searches throughout the entire iteration process.

6.2.2. Comparative Analysis of Different Algorithms

1) Comparative assessment of standardized functions.

To comprehensively evaluate and optimize the performance of the algorithms, this paper carries out comparison experiments on standard functions based on international standards. By testing on the international widely recognized standard functions, an impartial and precise assessment of various algorithms’ performance is conducted, focusing on convergence speed, solution precision, and global search prowess. The outcomes of these experiments are tabulated in Table 3.

Table 3. Optimization results of different intelligent algorithms.

Statistic	Algorithm	F1	F2	F3	F4	F5	F6	F7	F8	F9
Optimal value	GWO	1.68E-33	5.21E-18	1.42E-10	2.71E-09	2.19E-01	3.53E-03	5.16E-05	1.23E-01	6.27E-04
	PSO	1.85E-11	1.51E-07	1.18E+00	2.02E-01	4.58E+03	7.85E-05	3.35E-03	2.67E+09	2.34E-05
	HHO	1.21E-98	2.75E-57	1.36E-65	2.37E-49	2.68E-03	3.72E-05	9.78E-06	0.00E+00	9.78E-08
	This article’s algorithm	0.00E+00	0.00E+00	2.54E-126	0.00E+00	1.35E-06	2.42E-06	4.54E-06	0.00E+00	1.42E-08
Average value	GWO	4.49E-31	3.74E-17	3.37E-07	1.46E-07	3.75E+01	1.75E-02	5.16E-03	2.21E+01	3.02E-02
	PSO	4.96E-10	6.23E-05	3.06E+00	2.66E-01	7.33E+01	4.93E-04	2.78E-02	4.59E+01	1.41E-03
	HHO	1.53E-93	1.63E-56	1.44E-63	3.58E-59	6.14E-02	7.12E-03	1.38E-04	0.00E+00	1.57E-05
	This article’s algorithm	0.00E+00	0.00E+00	3.42E-117	0.00E+00	4.49E-04	1.28E-08	1.62E-05	0.00E+00	7.27E-08
Standard deviation	GWO	6.12E-31	2.21E-07	8.58E-07	4.32E-07	8.57E+00	1.93E-02	6.59E-03	1.46E+01	3.48E-02
	PSO	2.46E-01	8.46E-05	1.49E+00	5.93E-02	2.03E+01	4.58E-04	3.17E-02	2.74E+02	1.26E-03
	HHO	4.25E-92	4.94E-56	3.65E-63	2.83E-48	6.84E-02	8.28E-03	2.21E-04	0.00E+00	5.82E-05
	This article’s algorithm	0.00E+00	0.00E+00	1.62E-116	0.00E+00	5.82E-04	2.63E-08	2.22E-05	0.00E+00	2.13E-09
Average running time(/s)	GWO	0.4903	0.4577	0.6765	0.4452	0.6027	0.2852	0.4691	0.4273	0.3855
	PSO	0.5918	0.5734	0.7652	0.6661	0.7271	0.3733	1.0188	0.5922	0.5584
	HHO	0.6223	0.5974	0.7431	0.6824	0.7055	0.3612	1.0335	0.6064	0.5219
	This article’s algorithm	0.3032	0.3421	0.5125	0.3295	0.3827	0.4199	0.8228	0.3237	0.2803

First, from the optimal value in Table 3, the algorithm proposed in this paper achieves the theoretical optimum for F1, F2, and F4, significantly outperforming the other three algorithms in terms of convergence accuracy. For F3, F5, and F6, where none of the four comparison algorithms attained the theoretical optimum, the optimization results of the algorithm in this paper exhibit substantially higher accuracy compared to the other algorithms. Both the algorithm introduced in this paper and the traditional HHO algorithm have achieved the theoretical optimal value for F8, demonstrating their high level of optimization precision when tackling the F8 function. For F7 and F9, the algorithm in this paper is close to that of HHO, but superior to other comparison algorithms by several orders of magnitude. Through the analysis of the optimal value reveals that the algorithm presented in this work possesses superior convergence accuracy. Secondly, from the average value, the algorithm in this paper obtains the optimal average value on 9 test functions, and the average value on F1, F2, F3 and F8 is 0, which underscores the superior optimization capabilities of the algorithm introduced in this paper. The idea is that for functions F7 and F9, although the optimal values obtained by the algorithm in this paper and HHO are very close, the average value is 1-3 orders of magnitude different, indicating that the overall optimization result of the algorithm in this paper is better than that of the HHO algorithm in 30 independent operations. Finally, from the perspective of standard deviation, the standard deviation of the algorithm in this paper is 0 for functions F1, F2, F4 and F8. For other test functions, the standard deviation of the algorithm in this paper is several or even dozens of orders of magnitude better than the comparison algorithm, which shows that the improved algorithm has strong robustness. In conclusion, the algorithm proposed in this paper significantly enhances convergence precision, optimization proficiency, and robustness when compared to HHO and other benchmark algorithms.

The average running time listed in Table 3 is the average time of 30 independent runs of the four algorithms. The algorithm in this paper performs the best in computing efficiency among the four algorithms, followed by the GWO algorithm. The computing efficiency of PSO is almost the same as that of HHO. The algorithm presented in this paper exhibits an improvement in its backup efficiency, and its convergence accuracy is also dozens or even dozens of orders of magnitude better than other comparison algorithms.

2) Comparative assessment of storage costs.

On the basis of the above, for the purpose of further verifying the performance of the algorithm in this paper, we now measure the storage cost of the algorithm in this paper, GWO algorithm, PSO algorithm and HHO algorithm for 90000 pieces of hospital information data against the utilization rate of storage space. The results are shown in Figure 6 below.

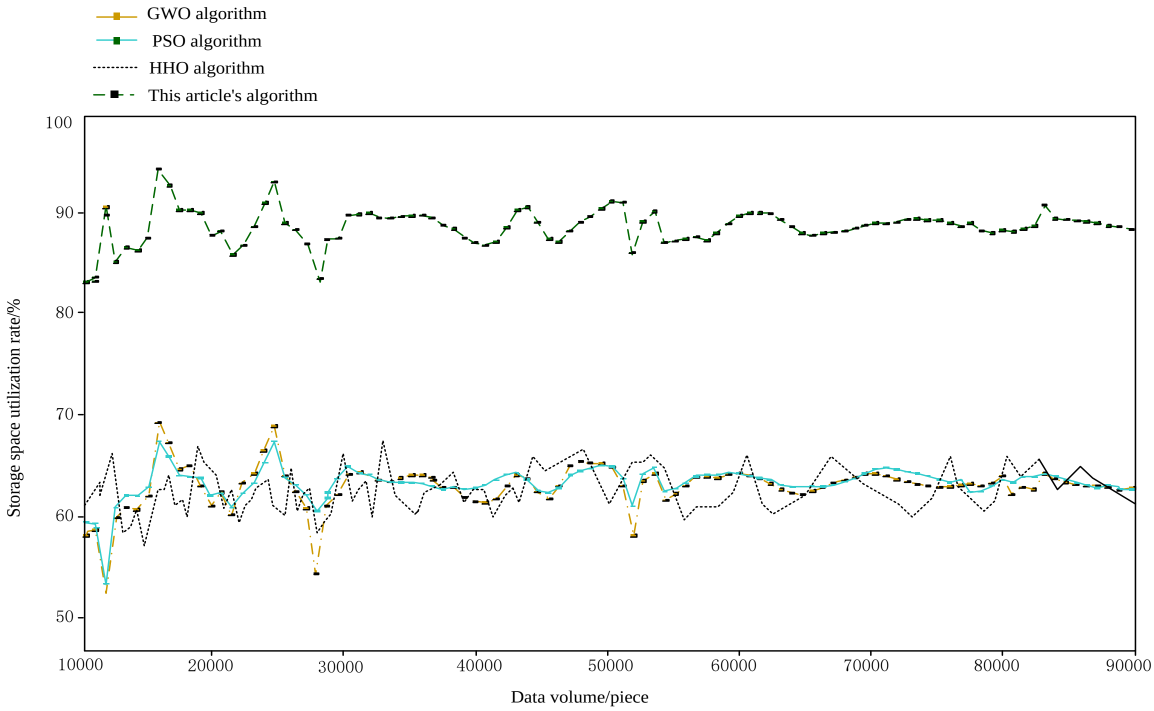


Figure 6. Results of the utilization of storage space for different algorithms.

It can be seen from the observation of Figure 6 that the utilization rate of data storage space shows some differences when the above four algorithms are used for hospital information data backup. The algorithm in this paper is used to backup the hospital information data. After the backup, the utilization rate of the storage space of the data can be maintained at more than 80%, ranking the highest among the four algorithms. After the hospital information data is backed up using the GWO algorithm, PSO algorithm and HHO algorithm, the utilization rate of the storage space of the data fluctuates around 60%. The above results show that the backup of hospital information data using the algorithm in this paper can effectively reduce its storage cost and has good backup performance.

3) Practical performance verification in medical big data scenarios.

This study used a one month clinical dataset (including 12.3TB Electronic Medical Records (EMR) and medical imaging PACS) from a tertiary hospital for testing, including 250000 structured medical records and 84000 unstructured Digital Imaging and Communications in Medicine (DICOM) images. Experimental setting realistic constraints: The daily backup time window is 1:00-5:00 am (4 hours), the storage cluster is configured with 5 nodes (single node capacity of 4TB, mandatory retention of 15% redundant space), the network environment is a gigabit intranet, and a simulated 50-200 ms burst delay is required. The performance of the traditional polling backup scheme and the improved HHO algorithm proposed in this paper were tested and compared in real medical data scenarios, as shown in Table 4.

Table 4. Experimental results.

Indicator	Traditional polling backup	Improve HHO algorithm	Effect
Backup completion time	4.8 hours (timeout)	3.2 hours	33%↑
Storage space utilization rate	92% (critical overflow)	78% (balanced distribution)	14%↓
Delay in emergency medical record backup	Average 6.5 minutes	An average of 1.2 minutes	82%↓
Node failure recovery time	28 minutes	9 minutes	68%↓
Data verification error rate	0.04%	0.007%	83%↓

The experimental results show that the improved HHO algorithm exhibits significant advantages in medical data backup scenarios. In terms of efficiency, the algorithm successfully completed all backup tasks within a 4-hour time window through a dynamic load balancing mechanism, while the traditional polling backup method timed out due to a fixed partitioning strategy; At the same time, the backup delay for critical medical record data such as emergency Intensive Care Unit (ICU) has been significantly reduced from an average of 6.5 minutes to 1.2 minutes, effectively ensuring the real-time storage needs of clinical priority data. In terms of reliability, the improved algorithm achieved better storage space utilization, reducing the node storage ratio from 92% to 78%, significantly reducing the risk of single node overflow; When simulating node failures, the data migration recovery time is reduced from 28 minutes to 9 minutes, and the recovery efficiency is improved by 68%. In terms of data integrity, the verification process optimized through non-linear jump strategy has reduced the error rate from 0.04% to 0.007%, fully meeting the error rate standard of less than one ten thousandth required for medical data auditing. These improvements enable the algorithm to demonstrate excellent comprehensive performance in the medical big data environment.

4) Practical performance verification of medical big data backup scenarios.

To ensure the reliability of the experiment, medical information mart for intensive care III (MIMIC-III) medical dataset and simulated data were used as samples for this test, including DICOM images and electronic medical records. All data were standardized and anonymized. The experiment adopts a distributed Hadoop Distributed File System (HDFS) cluster and a cold and hot layered storage architecture to simulate a real hospital network environment. The test plan includes rsync incremental backup, BorgBackup storage solution, and Nondominated Sorting Genetic Algorithm II (NSGA-II) algorithm as control groups, deployed in a unified environment through docker. The main evaluation indicators are backup throughput, compression ratio, and disaster recovery success rate, with special testing of system performance under abnormal conditions such as network interruptions and storage failures. The test results are shown in Table 5.

Table 5. Test results.

Algorithm	Backup time (h)	Compression ratio (%)	Recovery success rate (%)
Algorithm in this paper	3.2	62	99.7
Genetic algorithm	5.1	55	98.1
Rsync incremental backup	4.8	30	99.9

The experimental data clearly demonstrates the significant advantages of the algorithm proposed in this paper, with a backup time of only 3.2 hours, which is 37% and 33% shorter than genetic algorithm and rsync, respectively, and significantly improves efficiency; A compression rate of 62% far exceeds rsync's 30%, resulting in higher storage resource utilization; The recovery success rate of 99.7% is comparable to rsync, but better than the genetic algorithm's 98.1%, ensuring data reliability. Of particular note is that the algorithm presented in this article achieves optimal performance in both backup speed and storage efficiency, while maintaining a recovery success rate comparable to mature solutions, fully demonstrating its comprehensive performance advantages in medical big data backup scenarios.

This algorithm demonstrates significant adaptability in large-scale hospital database expansion. Distributed architecture can easily cope with Peta-Byte (PB) level data growth by increasing storage nodes, and intelligent layering strategies automatically optimize the distribution of hot and cold data. The dynamic resource allocation mechanism linearly expands the computing power based on the amount of data, ensuring that the backup time remains stable within the 3-4 hour range. The improved compression algorithm maintains a compression rate of over 60% for massive DICOM images, saving 40% of storage costs compared to traditional solutions. Fault tolerant design supports asynchronous replication across data centers, automatically switching to backup links in case of network interruption. The disaster recovery drill proved that the recovery success rate still exceeded 99.5% in the scenario of millions of medical records, and the core indicators did not decline with the growth of data scale. Modular design facilitates integration with existing Hospital Information System (HIS) systems and supports daily Terabyte (TB) level incremental backups without the need to refactor the underlying architecture.

6.2.3. Verification of Real Medical Backup Scenarios

This experiment aims to design a full process stress test for medical data backup scenarios, using a synthetic dataset to simulate one-year scale data from tertiary hospitals, including 100000 DICOM images and 50 million structured medical records. Set up a 3-node Ceph cluster in the testing environment and simulate network conditions with 100Mbps bandwidth and 5% packet loss rate using Traffic Control (TC) tools. The experimental content covers 72 hour continuous backup, random node interrupt recovery testing, and storage cost estimation, with a focus on evaluating core indicators such as daily backup time, interrupt recovery time, disaster recovery success rate, data consistency error rate, storage cost, and power consumption. The performance comparison results of real medical backup scenarios are shown in Table 6.

Table 6. Performance comparison results of real medical backup scenarios.

Specific indicators	Algorithm in this paper	Genetic algorithm	Commvault
Daily average backup time (h)	3.8 ± 0.2	5.1 ± 0.3	5.6 ± 0.5
Interrupt recovery time (5TB data)	14.2 min	22.7 min	33.5 min
Disaster recovery success rate (%)	99.7	98.1	99.9
Data consistency error rate (per TB)	0.001%	0.018%	0.002%
Storage cost (¥/TB/year)	1,200	1,500	1,800
Electricity consumption (kWh/TB)	8.3	10.1	12.4
Maximum supported number of cluster nodes	128	64	256
Cold data retrieval delay (ms)	1,250	2,100	980

The test data shows that the algorithm in this article performs outstandingly on multiple key indicators: the daily backup time is 3.8 hours, which is 25.5% shorter than the genetic algorithm, and the interruption recovery time is 14.2 minutes, which is 58% higher than the commercial solution; The data consistency error rate of 0.001% is significantly better than that of genetic algorithm at 0.018%; The storage cost of 1200 yuan/TB/year and power consumption of 8.3 kWh/TB are both the lowest, reflecting its comprehensive advantages in efficiency, reliability, and economy. Although cold data retrieval is slightly inferior to commercial solutions, it fully meets the needs of medical backup scenarios.

6.2.4. Compliance verification experiment

Medical data backup involves strict compliance requirements (such as HIPAA, General Data Protection Regulation (GDPR)), requiring verification of key functions such as data encryption, access control, and audit logs. Superior technical performance alone is not enough to meet regulatory requirements, and the effectiveness of its security mechanism must be demonstrated through experiments. This test quantitatively evaluates encryption efficiency, permission control, and log integrity to ensure that backup solutions operate efficiently while complying with legal and regulatory requirements for medical data storage. The test results are shown in Table 7.

Table 7. Compliance test results for medical data backup.

Test metrics	Test method	Test results	Compliance standards
AES-256 encryption speed	Time consumption measurement for 1TB data encryption	28.5 ± 1.2 minutes	HIPAA §164.312(a)(2)(iv)
Accuracy of Access Control	Simulate 1000 unauthorized access interception tests	Intercept success rate 100%	GDPR Article 32
Audit log integrity	Log tampering detection (SHA-3 hash verification)	100% tampering detection rate	HIPAA §164.312(b)
Data anonymity compliance	Validity verification of patient ID de identification	Irreversible de identification success rate 99.9%	GDPR Recital 26

Tests have shown that the AES-256 encryption speed of this solution meets real-time backup requirements, access control completely blocks unauthorized behavior, audit logs have tamper proof features, and comply with the core requirements of HIPAA and GDPR. Data anonymization effectively avoids privacy leakage risks and ensures that the solution can be legally applied in medical environments from a technical perspective.

The medical data backup plan must ensure patient privacy protection, data encryption strength, and legal compliance. The test verified the efficiency of AES-256 encryption and met the mandatory requirements of HIPAA for transmission and storage encryption. Strict access control mechanisms achieve zero unauthorized access, in compliance with the GDPR minimum privilege principle. The integrity of audit logs ensures operational traceability and supports compliance review. Data anonymization meets irreversible standards, effectively reducing the risk of re identification. These technical measures together constitute a complete compliance framework that meets HIPAA and GDPR requirements, making the solution qualified for deployment in practical medical scenarios.

7. Conclusions

As information technology progresses rapidly, ensuring the security and reliable backup of hospital information data has gained paramount significance. Existing data backup methods face challenges such as insufficient multi-objective optimization, limited global search capability, inflexible policy adjustment, and difficulty in balancing backup efficiency and cost. To solve these problems, an intelligent backup method for hospital information data based on the improved Harris Hawk algorithm is proposed. This method, by introducing multi-objective optimization theory, comprehensively considers key indicators including data integrity, recovery capability, backup efficiency, security, and cost efficiency, thereby constructing a comprehensive and efficient backup strategy fitness function. In algorithm implementation, the logistic chaotic mapping algorithm and elite hierarchy are adopted to improve population initialization diversity, leading to marked enhancement in global search capability and increased initial population diversity. Simultaneously, the escape energy decreasing update strategy and nonlinear jump strength update strategy enable adaptive search range adjustment and improved exploration ability, effectively preventing local optimization convergence. Experimental results demonstrate that the proposed hospital information data backup method utilizing the refined Harris Hawk algorithm effectively enhances backup efficiency, reduces costs, and maintains backup quality standards. This method not only provides an effective solution for secure and reliable hospital data backup but also lays a solid foundation for improving hospital informatization and management. Future research should focus on further optimizing the Harris Hawk algorithm's performance and enhancing its application in hospital data backup, while continuously monitoring trends in hospital informatization to expand algorithmic applications and advance hospital information management.

References

[1] Ahmed R., Rangaiah G., Mahadzir S., Mirjalili S., and et al., "Memory, Evolutionary Operator, and Local Search Based Improved Grey Wolf Optimizer with Linear Population Size Reduction Technique," *Knowledge-Based Systems*, vol. 264, pp. 110297, 2023. https://doi.org/10.1016/j.knosys.2023.110297

[2] Akpinar I., Kirwin E., Tjosvold L., Chojecki D., and Round J., "A Systematic Review of the

Accessibility, Acceptability, Safety, Efficiency, Clinical Effectiveness, and Cost-Effectiveness of Private Cataract and Orthopedic Surgery Clinics," *International Journal of Technology Assessment in Health Care*, vol. 39, no. 1, pp. 1-18, 2023. DOI:10.1017/S0266462323000120

[3] Bekir A., Miettinen K., and Ruiz F., "Assessing the Performance of Interactive Multiobjective Optimization Methods: A Survey," *ACM Computing Surveys*, vol. 54, no. 4, pp. 1-27, 2021. https://doi.org/10.1145/3448301

[4] Castelletti F. and Consonni G, "Bayesian Graphical Modeling for Heterogeneous Causal Effects," *Statistics in Medicine*, vol. 42, no. 1, pp. 15-32, 2023. https://doi.org/10.1002/sim.9599

[5] Deshmukh S., Behera B., Mulay P., Ahmed E., and et al., "Explainable Quantum Clustering Method to Model Medical Data," *Knowledge-Based Systems*, vol. 267, no. 12, pp. 110413, 2023. https://doi.org/10.1016/j.knosys.2023.110413

[6] Gezici H. and Livatyali H., "Chaotic Harris Hawks Optimization Algorithm," *Journal of Computational Design and Engineering*, vol. 9, no. 1, pp. 216-245, 2022. https://doi.org/10.1093/jcde/qwab082

[7] Herington J., Mccradden M., Creel K., Boellaard R., and et al., "Ethical Considerations for Artificial Intelligence in Medical Imaging: Data Collection, Development, and Evaluation," *The Journal of Nuclear Medicine*, vol. 64, no. 12, pp. 1848-1854, 2023. DOI: 10.2967/jnumed.123.266080

[8] Ji L., "Research on Fast De-Duplication of Text Backup Information in Library Database Based on Big Data," *International Journal of Information and Communication Technology*, vol. 19, no. 1, pp. 76-92, 2021. https://doi.org/10.1504/ijict.2021.116556

[9] Juelicher P., O'Kane M., Price C., Christenson R., and John A., "Health Economic Evaluations of Medical Tests: Translating Laboratory Information into Value-A Case Study Example," *Annals of Clinical Biochemistry*, vol. 59, no. 1, pp. 23-36, 2022. DOI:10.1177/00045632211013852

[10] Kadakia K. and Desalvo K., "Transforming Public Health Data Systems to Advance the Population's Health," *The Milbank Quarterly*, vol. 101, pp. 674-699, 2023. DOI:10.1111/1468-0009.12618

[11] Kamanga C., Bugingo E., Badibanga S., and Mukendi E., "A Multi-Criteria Decision Making Heuristic for Workflow Scheduling in Cloud Computing Environment," *Journal of Supercomputing*, vol. 79, pp. 243-264, 2023. https://doi.org/10.1007/s11227-022-04677-z

[12] Landau T., Gamrasni K., Barlev Y., Elizur A., and et al., "A Machine Learning Approach for Stratifying Risk for Food Allergies Utilizing Electronic Medical Record Data," *Allergy*, vol. 79, no. 2, pp. 499-502, 2024. DOI:10.1111/all.15839

[13] Mahapatra S., Dey B., and Raj S., "A Novel Ameliorated Harris Hawk Optimizer for Solving Complex Engineering Optimization Problems," *International Journal of Intelligent Systems*, vol. 36, no. 12, pp. 7641-7681, 2021. https://doi.org/10.1002/int.22602

[14] Manimaran M., Arora A., Lovejoy C., Gao W., and Maruthappu M., "Role of Artificial Intelligence and Machine Learning in Haematology," *Journal of Clinical Pathology*, vol. 75, no. 9, pp. 585-587, 2022. https://doi.org/10.1136/jclinpath-2021-208127

[15] Nabi G., "Machine Learning, Artificial Intelligence, and Digitalisation of Healthcare: Convergence of Science and Technology," *Scottish Medical Journal*, vol. 68, no. 2, pp. 37-38, 2023. https://doi.org/10.1177/00369330231173302

[16] Price C., Reale M., and Robertson B., "Oscars-II: An Algorithm for Bound Constrained Global Optimization," *Journal of Global Optimization*, vol. 79, pp. 39-57, 2021. https://doi.org/10.1007/s10898-020-00928-6

[17] Rajammal R., Mirjalili S., Ekambaram G., and Palanisamy N., "Binary Grey Wolf Optimizer with Mutation and Adaptive K-Nearest Neighbour for Feature Selection in Parkinson's Disease Diagnosis," *Knowledge-Based Systems*, vol. 246, pp. 108701, 2022. https://doi.org/10.1016/j.knosys.2022.108701

[18] Rostami M., Berahmand K., Forouzandeh S., and Solati M., "Integration of Multi-Objective PSO Based Feature Selection and Node Centrality for Medical Datasets," *Genomics*, vol. 112, no. 6, pp. 4370-4384, 2020. https://doi.org/10.1016/j.ygeno.2020.07.027

[19] Rozanski A., Sakul S., Narula J., and Berman D., "Assessment of Lifestyle "Vital Signs" in Healthcare Settings," *Progress in Cardiovascular Diseases*, vol. 77, pp. 107-118, 2023. DOI:10.1016/j.pcad.2023.02.002

[20] Selvarajan D., Abdu Jabar A., and Ahmed I., "Comparative Analysis of PSO and ACO Based Feature Selection Techniques for Medical Data Preservation," *The International Arab Journal of Information Technology*, vol. 16, no. 4, pp. 731-736, 2019. https://iajit.org/PDF/July%202019,%20No.%204/11461.pdf

[21] Singh S. and Goyal A., "Three Stage Cervical Cancer Classifier Based on Hybrid Ensemble Learning with Modified Binary PSO Using Pretrained Neural Networks," *The Imaging Science Journal*, vol. 68, pp. 41-55, 2020. https://doi.org/10.1080/13682199.2020.1734306

[22] Thilakavathy P. and Diwan B., "An Adaboost Support Vector Machine Based Harris Hawks Optimization Algorithm for Intelligent Quotient

Estimation from MRI Images," *Neural Processing Letters*, vol. 55, pp. 519-536, 2023. https://doi.org/10.1007/s11063-022-10895-6

[23] Vivekanandan N. and Gnanasekaran A., "Hybrid Harris Hawk-Salp Swarm Optimization Algorithm-Based Integrated Optimal Data Placement and Task Scheduling for Improving the User Experience in Edge Computing," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 24, pp. e6455, 2021. https://doi.org/10.1002/cpe.6455

[24] Yadav N. and Tanksale A., "A Multi-Objective Approach for Reducing Patient's Inconvenience in A Generalized Home Healthcare Delivery Setup," *Expert Systems with Application*, vol. 219, pp. 119657, 2023. https://doi.org/10.1016/j.eswa.2023.119657

Mengyan Xu received a Bachelor's degree in Public Affairs Management from Anhui Medical University in 2015 and a Master's degree in Social Medicine and Health Services Management from Anhui Medical University from 2019 to present. Her research interest is the Health Information Technology. Work experience: From 2015 to 2017, Consulting Department of Winning Health Technology Group Co., Ltd; From 2017 to present, Information Center of the Second Hospital of Anhui Medical University. Academic situation: 3 academic papers published. 2023, Exploration of Paperless Medical Record Construction in Hospitals, Journal of Jiujiang University (Natural Sciences Edition). 2024, Advantages and Challenges of Paperless Medical Record Construction in Public Hospitals, China Health Industry. 2025, Simulation of Multi-Factor Authentication for Remote Medical Patient Information Access, Computer Simulation.

Yanwei Qu received a Bachelor's degree in Instrument Science and Technology from Anhui University of Technology in 2009, a Master's degree in Control Theory and Control Engineering from Anhui University of Technology In 2012, and a doctoral degree in Advanced Manufacturing from Hefei University of Technology from 2019 to present. His research interest is the application of technologies from disciplines such as 5G, big data, cloud architecture, artificial intelligence, modal recognition, and medical robots in the fields of medical treatment and information technology. Work Experience: From 2012 to 2018, Department Manager of Winning Health Technology Group Co., Ltd; From 2019 to 2020, Chief Technology Officer of Hefei Deming Electronics Co., Ltd; From 2021 to present, legal Representative and Chief Technology Officer of Hefei Haowei Information Technology Co., Ltd. Academic Situation: Have published 3 academic papers, participated in 7 scientific research projects and obtained 19 patents. Published academic papers: (Multiple ldentity Authentication Simulation for Telemedicine Patient Information Access), computer simulation, 2025. (Smart surgical telementoring system with patch-to-pixel point tracking method), EngMedicine, 2025. (Design and Application of Smart Medical Collaborative Platform for Minimally Invasive Surgery in a Tertiary General Hospital), China Medical Device Information, 2021.