Multi-Objective Cold Chain Logistics Optimization via Integrated K-means Clustering and Genetic Algorithm

Weilu Yuan*, Jia Wang
Anyang Vocational and Technical College, Anyang 455000, Henan, China
E-mail: 18438659736@163.com
*Corresponding author

Keywords: K-means clustering, genetic algorithm, cold chain logistics, logistics delivery, path planning

Received: May 13, 2025

With the growing demand for fresh food, Cold Chain Logistics (CCL) faces challenges in meeting market demands. To reduce logistics costs and improve delivery efficiency, a cold chain distribution path planning model combining K-means algorithm and Genetic Algorithm (GA) has been proposed. The K-means algorithm is used for cluster analysis to generate initial candidate locations for distribution centers, while the GA determines optimal center positioning. By integrating the rapid convergence of K-means with the global search capability of GA, this approach resolves local optimization issues. Cluster effectiveness is evaluated using contour coefficients. Subsequently, a multi-objective optimization model incorporating real-time traffic conditions, product freshness preservation time, and vehicle load constraints is constructed. This model was validated using Google Open Routes Data (GORD) and Vehicle Routing Problem (VRPI) instance datasets. The results indicated that the fusion algorithm performed well in optimizing CCL distribution paths. The average task processing time of the fusion algorithm was controlled between 6.32 seconds and 8.42 seconds, with the lowest resource utilization rate of only 75.21% to 78.46%, and an average energy consumption value of 149.67 J to 160.72 J. The delivery cost and efficiency per kilometer were approximately 1.2 yuan and 40 km/h. The dynamic response capability of path planning has been significantly enhanced, effectively avoiding traffic congestion nodes and reducing cargo losses in cold chain transportation. This study has achieved collaborative optimization of distribution center location and path planning, which is of great significance for reducing operating costs, improving distribution efficiency, and promoting the construction of smart logistics systems.

Povzetek: Model K-means + genetski algoritem optimizira lokacije centrov in poti hladne verige ter zmanjša stroške in izgube z boljšim izogibanjem prometnim zastojem.

1 Introduction

The progress of society has led to a continuous improvement in people's quality of life. Online shopping and home delivery services have become a common way of life for modern residents [1]. Due to the increasing need for food, Cold Chain Transportation (CCT) is becoming a vital industry in modern times [2]. In today's era, with the continuous increase of population, people have a higher demand for fresh food, thus requiring more efficient Cold Chain Logistics (CCL) systems [3]. The current CCL technology is relatively backward, and the CCL cannot meet the requirements of consumers [4]. The overall distribution system is still short of cold storage and manpower support, causing higher expenditures for the CCL system [5]. Many scholars have researched CCL path planning. Bai et al. considered the complexity of road networks and time-varying traffic conditions, studied the low-carbon vehicle routing problem of CCL, and proposed a corresponding model. Compared to traditional methods, it could more accurately evaluate delivery costs and Carbon Emissions (CE) [6]. Pu et al. incorporated driving distance and actual loading capacity into the multi-warehouse vehicle routing problem and proposed a mixed integer programming model. Under this model, CE could be reduced by traveling shorter distances, providing methodological guidance for vehicle route planning in terms of logistics costs and CE [7]. Yin proposed an improved non-dominated sorting Genetic Algorithm (GA) solution model based on the vehicle routing problem, while meeting the customer's cargo and time requirements. The path strategy calculated by this model had significant value in reducing CE while meeting customer demand for goods and time requirements [8]. Du et al. found that existing express delivery companies are facing the challenge of improving customer satisfaction while ensuring total cost control. Therefore, researchers have proposed a new multi-objective solution model for joint distribution. This model could effectively reduce costs and CE while guaranteeing higher customer satisfaction [9]. Li et al. put forth an improved model that combines the Carbon Trading (COT) mechanism to solve the position path inventory problem. After optimizing the location costs, transportation, inventory, etc., it was found

that the model could effectively lower CE in the COT environment, but the major environmental element for cost reduction was CCL [10].

The K-means algorithm has become a classic tool in clustering analysis due to its simplicity, efficiency, and wide applicability. Ashari et al. used the K-means algorithm for classification analysis and the Rand method for optimization to determine the classification of flood-affected areas in some regions. This method could quickly perform cluster analysis [11]. Şenol proposed a fusion K-Means method based on multivariate kernel density estimation to determine the optimal initial centroid for clusters generated by K-means, which was affected by randomly selected initial centroids. This algorithm used kernel density estimation to find the best initial centroid for identifying density regions, enabling fast and accurate clustering while reducing the impact of initial centroids [12]. The principle of GA is simple, it has strong global search capability and is easy to combine with other algorithms. Fu et al. proposed a PSO genetic hybrid algorithm built on phagocytosis to effectively schedule massive tasks in cloud environments. This algorithm ensured the particle diversity in the population, reduced the possibility of getting stuck in local optima, significantly shortened all the completion times of cloud tasks, and had high precision [13]. Choudhury et al. put forward a method that combines GA and detects fake news on different datasets. This method achieved the highest accuracy of 97% on the fake recruitment dataset [14]. Abedpour et al. developed a fusion method that combines GA and K-means clustering methods to lower clustering errors and optimize the overall performance of the network system. Firstly, it utilized K-means for clustering, and then used GA to allocate resources to devices with the Minimum Error Rate (MER). Experiments have shown that this method helps minimize latency and outperforms other algorithms in the objective function and MER [15]. The summary and comparison of the literature are shown in Table 1.

Table 1: Summary of literature content

Reference	Method	Optimization Objectives	Limitations	This paper improves
Bai et al. [6]	Low-carbon VRP model	Delivery cost, CE	Ignored perishability & location-routing integration	Integrated site-path optimization+cargo loss constraint
Pu et al. [7]	Mixed integer programming	CE reduction, Logistics cost	No real-time traffic adaptation	Real-time traffic constraint+adaptive genetic crossover
Yin [8]	Improved NSGA-III	CE reduction, Customer demand & time windows	Single-objective focus; No location optimization	Multi-target site selection+path collaborative optimization
Du et al. [9]	Multi-objective joint delivery	Cost, CE, Customer satisfaction	Static constraints; No energy optimization	Dynamic transportation+energy consumption minimization
Li et al. [10]	Carbon-trading mechanism	Location cost, Transport, Inventory, CE	High computational complexity	K-means initialization accelerates convergence
Ashari et al. [11]	K-means + Rand method	Flood area classification	Not applied to logistics	K-means is adapted to logistics node clustering
Şenol [12]	K-means + Kernel density	Cluster accuracy, Initial centroid optimization	No integration with routing algorithms	K-means output is seamlessly integrated with GA routing
Fu et al. [13]	PSO-GA hybrid	Task scheduling efficiency, Completion time	Not tailored for logistics constraints.	Customized cold chain perishability+vehicle load constraints

In summary, the current CCL faces challenges in meeting the growing market demand for fresh food products, primarily due to high costs, low efficiency, severe cargo loss, and excessive CE. The existing methods for locating and routing cold chain Distribution Centers (DCs) exhibit limitations, such as insufficient global search capability, slow convergence speed, and limited handling of multidimensional constraints, making them unsuitable for handling complex real-world scenarios. To achieve multi-objective collaborative optimization in cold chain DC location and route planning, this study innovatively proposes a hybrid model integrating K-means clustering with GA. The study employs K-means for initial candidate center generation and combines GA's global search capabilities to determine optimal DC locations. The clustering effectiveness is evaluated using silhouette coefficients, and a multi-objective optimization model is constructed to comprehensively consider real-time traffic conditions, product freshness requirements, and vehicle load restrictions. Experimental validation through Google Open Routes Data (GORD) and Vehicle Routing Problem (VRPI) datasets demonstrates the

model's effectiveness in optimizing CCL routes. The study aims to reduce distribution costs, enhance efficiency, and minimize energy consumption, providing scientific decision support for optimal site selection. This approach expands industry applications and promotes intelligent logistics system development while significantly reducing CE, advancing sustainable growth, and environmental protection in the logistics sector.

2 Methods and materials

2.1 Design of CCL-DCL model based on K-means algorithm

The core goal of CCL-DCL is to achieve multi-objective collaborative optimization of construction cost, transportation efficiency, service quality, and environmental sustainability [16]. However, CCL's DCL faces challenges such as difficulty in accurately predicting demand fluctuations, high construction and operational costs, insufficient data, and limited technical support. The K-means has become a classic tool in clustering analysis since its simplicity, efficiency, and wide applicability. Therefore, this study uses K-means to perform DCL on CCL. The Euclidean distance calculation in K-means is given by equation (1).

$$d(X,C_i)=\sqrt{\sum_{j=1}^{m}\left(X_j-C_{ij}\right)^2} \tag{1}$$

In equation (1), C_{ij} is the j-th attribute of C_i. C_i is the i-th CC point. X_j is the j-th attribute of X. X is a data entity. m is a measure of data objects. The evaluation of clustering effect often uses contour coefficient as a standard, as shown in formula (2).

$$S^{(i)}=\frac{b^{(i)}-a^{(i)}}{\max\left\{b^{(i)},a^{(i)}\right\}} \tag{2}$$

In equation (2), $a^{(i)}$ and $b^{(i)}$ are the average distances between the remaining sample points and all other sample points within the adjacent cluster. The contour coefficient when $S^{(i)}$ represents i. The workflow of K-means is displayed in Figure 1.

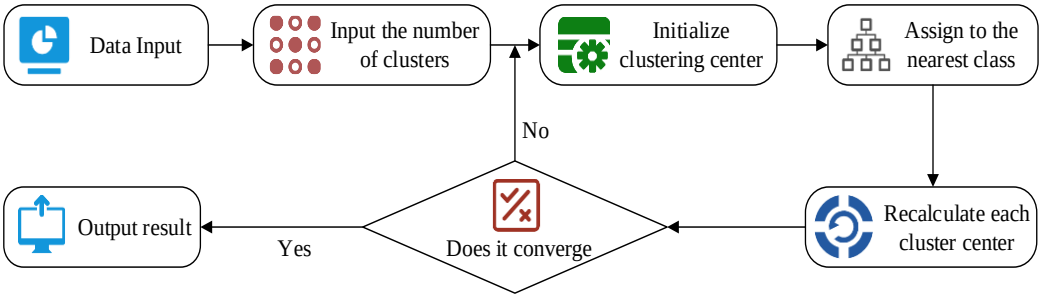


Figure 1: K-means algorithm workflow diagram

In Figure 1, K-means first measures the distance between all samples and the Cluster Center (CC) based on Euclidean distance, and divides each sample into clusters corresponding to the centroid of the nearest neighbor. Subsequently, it recalculates the centroid coordinates based on the current cluster members and defines it as the geometric mean of the samples within the cluster. The system iteratively performs sample allocation and centroid update operations until the pre-centroid coordinates are stable or the maximum Number of Iterations (NoI) is reached, thereby achieving dynamic optimization-based clustering partitioning. Finally, K centroids and the data allocation results for each class are output [16, 17]. The fixed cost calculation consisting of land leasing and personnel salaries is shown in equation (3).

$$F_1=\sum_{j=1}^{m}R_jZ_j \tag{3}$$

In equation (3), R_j is the unit price of land leasing or the salary of personnel per unit area. Z_j is the decision variable. The operating cost calculation of CCL DC is shown in equation (4).

$$F_2=\sum_{j=1}^{m}R_j^{(1)}\left(\sum_{i=1}^{l}X_{ij}B_{ij}Z_j\right)^{P'} \tag{4}$$

In equation (4), $R_j^{(1)}$ is the daily operating cost of the logistics hub. X_{ij} is the transportation volume of goods from the supplier to the DC. B_{ij} is the unit transportation cost coefficient between the supplier and the DC. p is the scale adjustment parameter. The transportation cost of the entire supply chain process is calculated. The total logistics cost covering the supply end to the cold chain transit node and then distributed to the end consumption node is shown in equation (5).

$$F_3=\sum_{i=1}^{l}\sum_{j=1}^{m}C_1X_{ij}d_{ij}B_{ij}Z_j+\sum_{j=1}^{m}\sum_{k=1}^{n}C_1X_{jk}d_{jk}Z_jB_{jk} \tag{5}$$

In equation (5), C_1 is the unit distance transportation cost. B_{jk} is the path complexity coefficient from the DC to the demand point. d_{ij} is the length of the path from the supplier to the transit base. X_{jk} is the transportation volume of materials from the supply end to the logistics node. d_{jk} is the average path length from the transit base to the consumer terminal. The overall spending of goods damage is given by equation (6).

$$F_4=C_2\sum_{i=1}^{l}\sum_{j=1}^{m}X_{ij}\left(1-e^{\frac{-\theta d_{ij}B_{ij}Z_j}{v}}\right)+C_2\sum_{j=1}^{m}\sum_{k=1}^{n}X_{jk}\left(1-e^{\frac{-\theta d_{jk}B_{jk}Z_j}{v}}\right) \tag{6}$$

In equation (6), C_2 is the unit cost of goods damage. θ is the coefficient of cargo damage rate. v is the transportation speed. To ensure constant temperature in the carriage, a significant amount of refrigeration costs will be incurred during transportation, as shown in equation (7).

$$F_5 = \sum_{i=1}^{l}\sum_{j=1}^{m} C_2 X_{ij} d_{ij} Z_j + \sum_{j=1}^{m}\sum_{k=1}^{n} C_2 X_{jk} d_{jk} Z_j B_{jk} \quad (7)$$

CCL refers to a logistics network that maintains a low temperature throughout the entire process based on the characteristics of the transported goods, to maintain the quality of the goods from production to delivery to consumers. The CCL delivery process is shown in Figure 2.

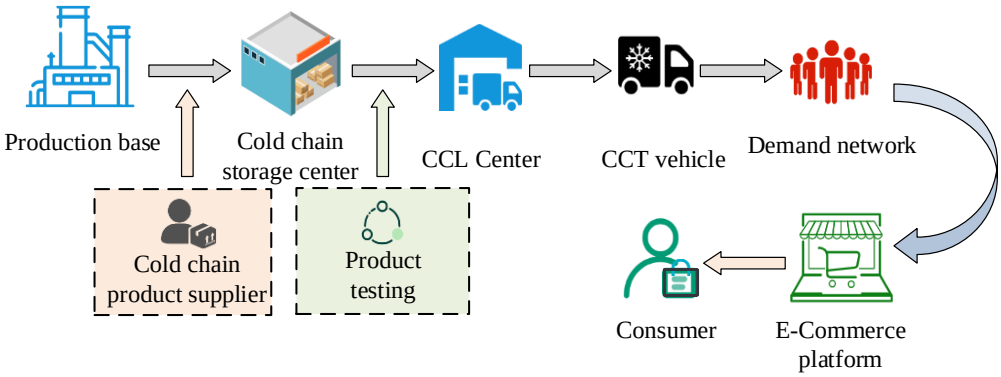


Figure 2: CCL distribution process

In Figure 2, the operation process of CCL starts from product information display and production, and involves product production and primary processing at the production base. After pre-cooling, processing, and classification, the product enters the cold chain storage center for refrigeration and preservation. Subsequently, quality inspection agencies conduct quality checks on secondary products to ensure compliance with standards. Third-party logistics companies transport products from storage centers to CCL DCs. The DC delivers products to demand outlets through cold chain transport vehicles. Finally, consumers purchase cold chain items through e-commerce platforms. Various kinds of CCL products have different characteristics and require different temperatures to be kept during CCT. As a result, to avoid losses caused by unbefitting temperature, standardized and unified operations can be carried out. Classifying the CCL products and suitable temperature is exhibited in Figure 3.

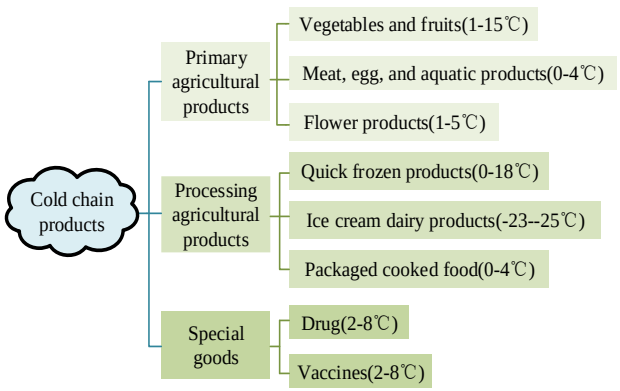


Figure 3: Classification of common cold chain items and their suitable temperatures

Figure 3 shows the ideal storage temperature range for certain specific products. This provides an important reference for ensuring the best quality of the product. However, in addition to paying attention to these fixed suitable temperatures, it is also necessary to closely monitor temperature control at various phases of transportation. Changes in environmental conditions during the handover of items may cause fluctuations in the surrounding temperature. Therefore, to ensure that the quality of the product is not affected, the temperature rise limit during transportation must be strictly controlled below -15 ℃. In addition, before loading the transport vehicle, it is necessary to ensure that the carriage has been pre-cooled to a temperature of -10 ℃ or lower to maintain a suitable temperature environment for the product throughout the entire transportation process.

2.2 Optimization of site selection model integrating K-means clustering and GA

K-means has efficient data processing capabilities and good adaptability for CCL-DCL, but is very sensitive to the selection of initial centroids and performs poorly in handling dynamic data or scenarios that require real-time adjustments [18]. The principle of genetic algorithm is simple, easy to implement and combine with other algorithms, with strong scalability and global search ability, which can effectively solve various types of problems. Therefore, this study integrates K-means with GA. The fast convergence ability of K-means can accelerate the search process of GA, while the global search ability of GA can compensate for the local optima of K-means. In addition, the robustness and flexibility of GA can enhance the performance of K-means in complex

data distributions. The GA operation process is shown in Figure 4.

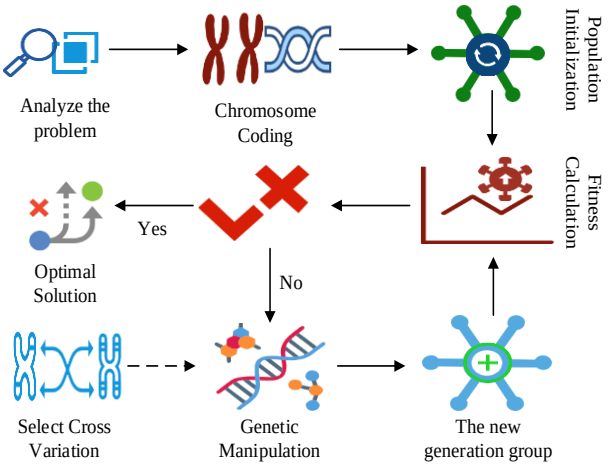


Figure 4: GS operation process

In the GA operation of Figure 4, the first step is to construct an algorithm parameter system. Based on decision vector features, the gene encoding dimension is determined, and chromosome sequence construction and population initialization configuration are performed. The next step is to activate the fitness evaluation mechanism, by establishing a nonlinear mapping model between the objective function and fitness values, to complete the quantitative evaluation of individual fitness. In the evolutionary stage, an optimized population is generated through selection operators, crossover recombination, and mutation mechanisms, and fitness evaluation and genetic operations are iteratively performed until the termination criteria of evolution are met. After the algorithm terminates, it is necessary to perform reverse decoding on the optimal gene sequence and output the final optimization solution [19]. This study takes the total cost of CCL as the determining factor for site selection, and constructs the objective function as shown in equation (8).

$$\min F = F_1 + F_2 + F_3 + F_4 + F_5 \tag{8}$$

In equation (8), F is the total cost of CCL. The formula for setting transportation convenience restrictions for site selection is shown in equation (9).

$$\mathrm{A}\left(\hat{C}_j\right) .. A_{\min}, \forall j \tag{9}$$

In equation (9), $\mathrm{A}\left(\hat{C}_j\right)$ denotes the analysis of the mean traffic convenience characteristics of service nodes within the radiation area $\hat{C}_j$ after the cold chain facility is deployed in the backup location. $A_{\min}$ is the minimum critical value of transportation convenience required for the effective operation of the cold chain distribution system. The constraint on CCT time is shown in equation (10).

$$\hat{t}_{j,l}, \hat{t}_{\max}, \forall j, l \tag{10}$$

In equation (10), $\hat{t}_{\max}$ means the maximum transportation time. The constraint on the spoilage of transported products is shown in equation (11).

$$0, \aleph(\hat{t}_{j,l} + \hat{t}_{i,j}) < \tilde{T}_{\max} \tag{11}$$

In equation (11), $\hat{t}_{i,j}$ is the transportation time between the production base and the CCL center. $\tilde{T}_{\max}$ is the maximum allowable transportation time limit for cold chain products to maintain freshness. The constraint on the supply-demand matching relationship of CCL is shown in equation (12).

$$\sum_{j=0}^{n'} h_{i,j} < C_j', \sum_{l \in L'} \hat{h}_j .. \hat{Q} \tag{12}$$

In equation (12), L' is the set of CCL centers that can serve demand point j. Chromosome encoding is the primary step in applying GA. Encoding is the process of mapping the solution state space. Figure 5 shows the chromosome coding diagram.

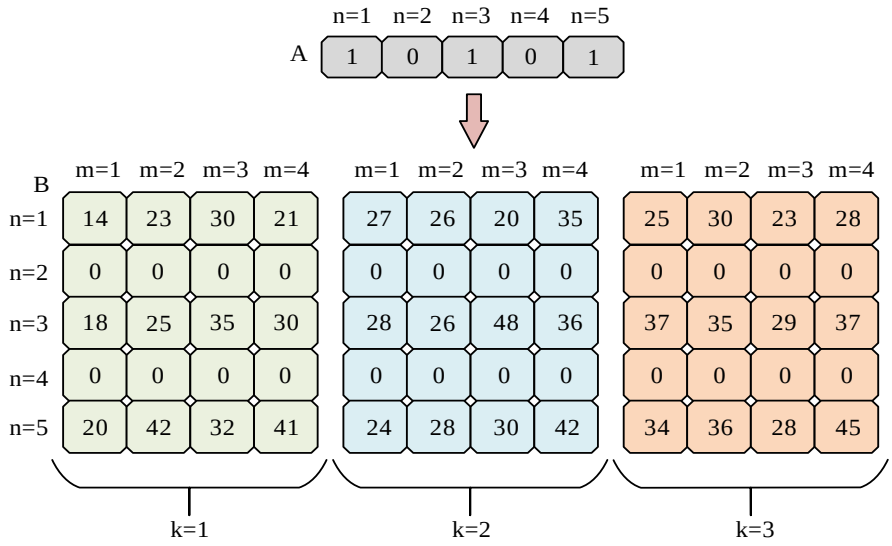


Figure 5: Chromosome coding diagram

In Figure 5, the chromosome encoding diagram illustrates how to select different n values through binary encoding of chromosome X, and displays the detailed data corresponding to these n values in matrix Y. Among them, chromosome X is a binary sequence with a length of 5, indicating whether to select certain values from n=1 to n=5. Matrix Y is a 2D table with rows representing different values of n. Columns represent different values of m, and the numbers in each cell represent a certain indicator or data. The columns of matrix Y are grouped into k=1, k=2, and k=3, with each group containing multiple values of m. In GA, the fitness function is a standard utilized to determine the chromosomes quality of the GA population. The corresponding fitness function selected is given by equation (13).

$$fit(i) = \frac{1}{F(i) + \mu} \tag{13}$$

In equation (13), μ is a positive real number within the interval $(0,1)$. $F(i)$ denotes the total cost function value of genetic unit i. $fit(i)$ is the fitness of i. The μ determined by the research through many experiments is set to 0.1. When μ varies between 0.05 and 0.2, the performance trend of the algorithm remains consistent. When μ is 0.1, the best balance can be achieved between cost minimization and algorithm stability. The strategy selection section constructs a probability function for elimination based on individual fitness, as shown in equation (14).

$$P_x = \frac{1 - f_x}{\sum \tilde{n}_{x=1} f_x} \tag{14}$$

In equation (14), P_x is the probability of being selected. $\tilde{n}$ is the population size. f_x is individual fitness. The genetic recombination mechanism achieves offspring generation by exchanging homologous segments of parental gene sequences. This process involves the localization of random breakpoints and the exchange of genetic material following Mendel's laws. To improve the adaptability of algorithms in processing and analysis, it is necessary to automatically adjust and optimize their processing methods, order, parameter settings, boundaries, and constraints based on the statistical distribution and structural characteristics of data [20]. The study introduces a two-point crossover algorithm. Two point crossover refers to randomly selecting two crossover points between two parental individuals, and then exchanging gene fragments between these two crossover points to produce two new offspring individuals. The expression is shown in equation (15).

$$\begin{aligned} C1_k &= \begin{cases} P1_k & k \le u \\ P2_k & \mathrm{u} < k \le s \\ P1_k & k > s \end{cases} \\ C2_k &= \begin{cases} P2_k & k \le u \\ P1_k & \mathrm{u} < k \le s \\ P2_k & k > s \end{cases} \end{aligned} \tag{15}$$

In equation (15), $C1$ and $C2$ represent the offspring chromosomes. $P1$ and $P2$ represent the parent chromosomes. u and s represent the crossover points. k represents the gene position index, where $k = 1, 2, 3 \ldots\ldots n$. The genetic crossover operation diagram is shown in Figure 6.

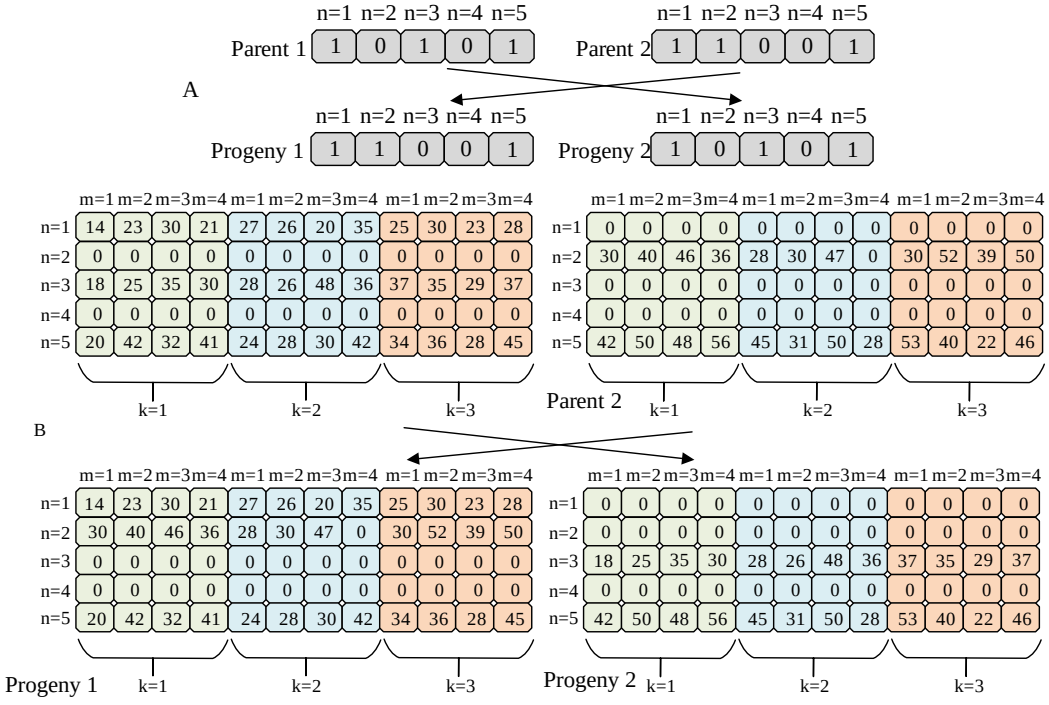


Figure 6: Schematic diagram of genetic crossover operation

In Figure 6, taking the parent individuals P_1 and P_2 as an example, the specific implementation of their crossover process is as follows: Firstly, an adaptive crossover probability model is constructed, which dynamically selects the second and third gene loci as crossover boundaries based on the current iteration times and population fitness variance. The next step is to perform gene fragment exchange - P_1 retains the first 2 genes and inherits the last 3 genes of P_2 to generate offspring C_1, while P_2 retains the first 2 genes and inherits the last 3 genes of P_1 to generate offspring C_2. In response to the possible phenomenon of gene homogenization in C_2, a gene position reversal protection mechanism will be triggered, and the final output will be a corrected offspring C_2. In summary, a crossover process involves two parent individuals and their offspring individuals. The specific crossover details are generated by selecting gene fragments from the parent individuals and combining them to create new offspring individuals. The fusion process of K-means and GA is shown in Figure 7.

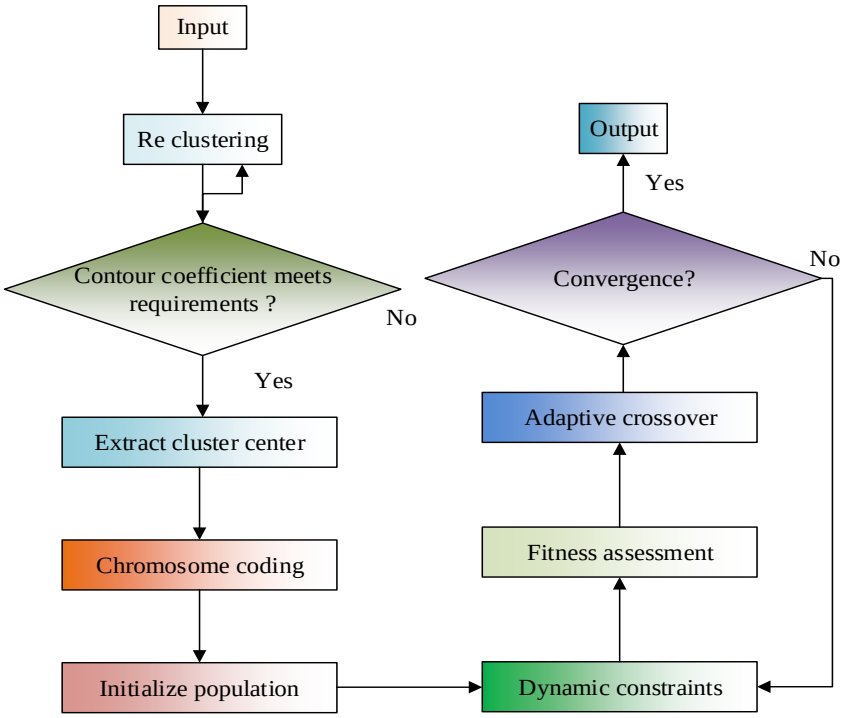


Figure 7: The fusion process of K-means and GA

As shown in Figure 7, the integration process of K-means and GA works as follows: First, the coordinates of demand nodes are input. Then, K-means clustering is performed and the k value is continuously adjusted until the contour coefficients met the requirements, at which point the cluster centers are extracted. Next, binary encoding is applied to the cluster center coordinates to initialize the GA population. The next step is to perform an iterative process, evaluate fitness, and check if clustering constraints are met. If satisfied, the total cost and dynamic constraints need to be calculated. Subsequently, adaptive crossover and mutation operations will be performed to determine convergence. If it does not converge, it will return to fitness evaluation for further iteration; If converged, it will output the optimal position and path.

3 Results

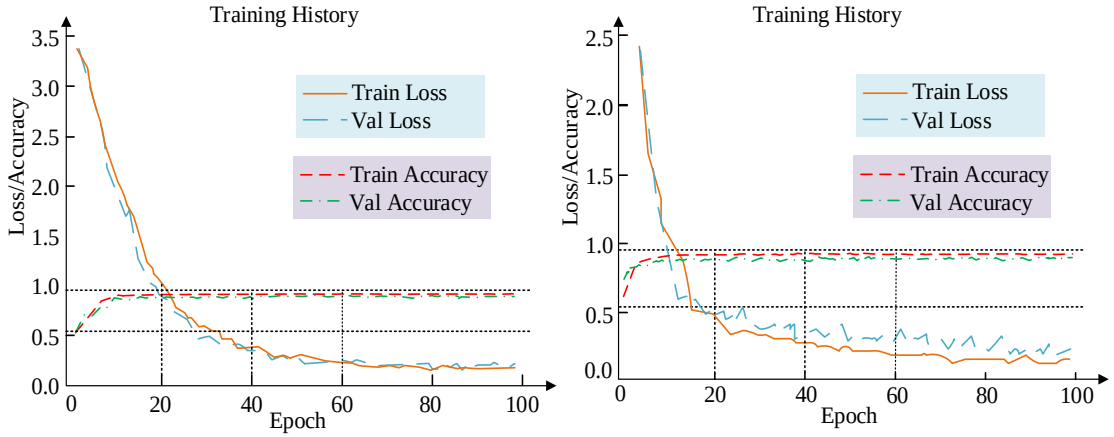
3.1 Performance testing of CCL delivery based on a new fusion algorithm

The experimental environment includes a CPU (Intel Core i9-11900K), GPU (NVIDIA RTX 3090), and 64GB DDR4 3200MHz memory. The operating system is Ubuntu 20.04 LTS, and the development framework is based on Python 3.8. The core logic of the algorithm is implemented through NumPy and SciPy. Path optimization parallel computing is completed by PyTorch. The study utilizes the GORD dataset and VRPI instance datasets to conduct detailed analysis of sample data, thereby validating the model's effectiveness and practicality. 500 sample data points are selected from the GORD dataset and 300 sample data points are selected from the VRPI dataset for the experiment. These samples encompass key characteristics including urban/suburban traffic conditions, node distribution, and delivery demands, effectively simulating real-world operational scenarios for CCL DCs. Specifically, GORD provides detailed information about actual urban road networks, such as road lengths, travel times, and congestion indices. VRPI contains critical parameters like node coordinates, delivery requirements, and vehicle load constraints. By comprehensively analyzing and processing these datasets, the performance of the model can be evaluated under various operating conditions. The advantages of data sets are shown in Table 2.

Table 2: Data set comparison analysis

Dataset Name	Geographic Coverage	Traffic Dynamic Characteristics	Data Authenticity	Applicability	Advantages
GORD	Urban and suburban areas	Includes road lengths, travel times, congestion coefficients, etc.	Real - world urban road network data	Suitable for urban and suburban distribution scenarios	Provide detailed real urban road network information, can accurately simulate the actual traffic conditions
VRPI	Diverse scenarios	Includes node coordinates, delivery demands, vehicle capacity constraints, etc.	Classic problem instances	Suitable for classic VRPI research	Offers rich VRPI instances with key parameters for easy model validation and comparison
Other data sets	Geographical coverage varies	The characteristics are not equal	Authenticity varies	The applicability is not equal	Specific information needed for CCL may not be available

In Table 2, the dataset used is more suitable for CCL distribution path planning than other datasets. The average runtime of the research algorithm is 120 seconds on the GORD dataset and 150 seconds on the VRPI dataset. To ensure reproducibility, the GA parameters are set as follows: population size (100), crossover rate (0.8), mutation rate (0.1), roulette wheel selection method, and maximum iteration count (500). For K-means clustering, the number of clusters is set to 5, initial centroids are randomly determined, and the maximum iteration count is 100. When the change in the centroid position is less than 0.001, it is considered that convergence has been achieved. This study first conducts a comparative analysis of the loss function and accuracy of the fusion algorithm for different delivery distances, as shown in Figure 8.



(a) Long distance delivery (b) Close range delivery

Figure 8: Comparison of accuracy and loss rate at different distances

According to Figure 8 (a), as the NoI increases, the loss value continuously decreases and the accuracy continuously increases. At the 100th iteration, the loss value reaches 0.06 and the accuracy reaches 99.89%. In Figure 8 (b), as the NoI increases, the algorithm performance improves, and the initial loss value rapidly decreases, ultimately reaching a stable state. The accuracy improves relatively quickly in the early stage, and gradually stabilizes later on. At the 100th iteration, the loss value reaches 0.08 and the accuracy reaches 99.67%. In summary, the research algorithm has a high calculation accuracy and low error loss for both long-distance and close-range delivery. This study tests and compares the convergence speed and energy consumption of the fusion algorithm, the Handling Multiple Objectives with PSO (MOPSO) algorithm, and the Multi-Objective Optimization algorithm (NSGA-III), as shown in Figure 9.

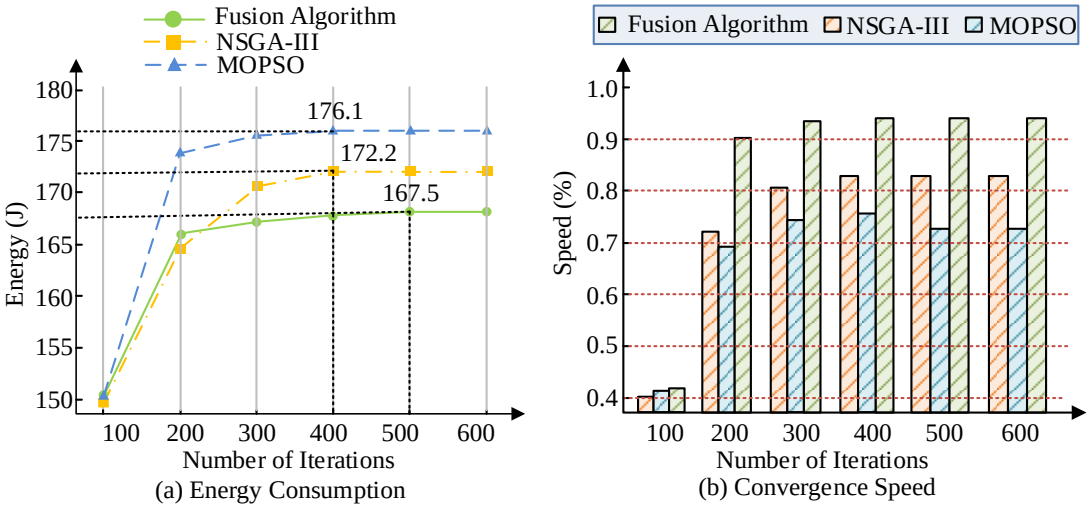


Figure 9: Convergence speed and energy consumption of different methods

In Figure 9 (a), MOPSO achieves a maximum energy consumption of 176.1 J after 400 iterations. NSGA-III has the highest energy consumption at 400 iterations, reaching 172.2 J. The fusion algorithm reaches its highest energy consumption of 167.5 J after 500 iterations, which is the lowest compared to the other two methods. In Figure 9 (b), the convergence speed of the fusion algorithm is the fastest, while the MOPSO method is the slowest. At 400 iterations, the fusion algorithm, NSGA-III, and MOPSO reach their highest points of 94.8%, 83.2%, and 76.3%, and then stabilize. This indicates that using fusion algorithms results in better energy consumption and convergence speed. This study tests the processing performance of fusion algorithms, NSGA-III, and MOPSO methods in GORD and VRPI, as listed in Table 3.

Table 3: The index test results of different multi-objective algorithms

Data set	Method	Average processing time/s	Resource occupancy rate/%	Average energy consumption/J
GORD	Fusion Algorithm	6.32	75.21	149.67
	NSGA-III	7.68	78.95	155.82
	MOPSO	8.95	82.36	162.47
VRPI	Fusion Algorithm	8.42	78.46	160.72
	NSGA-III	9.16	81.47	167.53
	MOPSO	10.43	84.62	174.29

In Table 3, the average task processing time of the fusion algorithm is controlled between 6.32 seconds and 8.42 seconds, which shows significant optimization effects and higher processing efficiency compared to other methods. Meanwhile, the fusion algorithm has the lowest computational resource utilization rate, ranging from 75.21% to 78.46%, indicating a higher efficiency in resource utilization. The average energy consumption value of the fusion algorithm is 149.67 J to 160.72 J, significantly lower than NSGA-III and MOPSO, and the energy optimization effect is significant. In summary, the fusion algorithm improves processing efficiency and reduces energy consumption and resource utilization through effective path planning. Its superior performance verifies its progressiveness and practical application value in CCL distribution path optimization. The study uses the t-test to analyze the treatment time, while the energy consumption data are evaluated by the Wilcoxon test, as shown in Table 4.

Table 4: Results of t-test and Wilcoxon test

Data set	Method	Average processing time/s	T test	Average energy consumption/J	Wilcoxon test
GORD	Fusion Algorithm	6.32 ±0.41	-	149.67±8.2	-
	NSGA-III	7.68 ±0.53	t=8.71, p<0.001, Δ=1.36 [1.02–1.70]	155.82±9.6	Z=3.89, p<0.002, Δ=6.15 [3.21–9.09]
	MOPSO	8.95 ±0.62	t=12.94, p<0.001, Δ=2.63 [2.18–3.08]	162.47±10.3	Z=5.32, p<0.001, Δ=12.80 [9.45–16.15]
VRPI	Fusion Algorithm	8.42 ±0.49	-	160.72±9.1	-
	NSGA-III	9.16 ±0.58	t=7.23, p<0.001, Δ=0.74 [0.51–0.97]	167.53±10.2	Z=3.12, p<0.008, Δ=6.81 [3.95–9.67]
	MOPSO	10.43 ±0.71	t=11.87, p<0.001, Δ=2.01 [1.62–2.40]	174.29±11.5	Z=4.95, p<0.001, Δ=13.57 [10.20–16.94]

In Table 4, the paired t-test results indicate that compared with NSGA-III, the average processing time of the research algorithm is reduced by 1.36 s, 2.63 s on GORD compared to MOPSO, and 0.74 s and 2.01 s on VRPI compared to MOPSO. Wilcoxon tests further reveal that the energy consumption of the proposed algorithm decreases by 6.15 J and 12.80 J on the GORD dataset, as well as 6.81 J and 13.57 J on the VRPI dataset. These findings confirm that the proposed fusion algorithm demonstrates statistically significant advantages in both computational speed and energy efficiency.

3.2 Application testing of logistics distribution based on new fusion algorithm

To validate the practical application effect of the new operation path mechanism, this study takes a certain fresh logistics enterprise as the research object. The cold chain delivery process is concentrated in the early morning hours, effectively avoiding the interference of traffic congestion on the experimental results. The experiment uses four trucks equipped with refrigeration units and cold storage insulation boxes to construct analysis samples based on actual distribution data from the enterprise. The comparison of the total delivery mileage between the fusion algorithm, NSGA-III, and MOPSO methods in urban and suburban areas is shown in Figure 10.

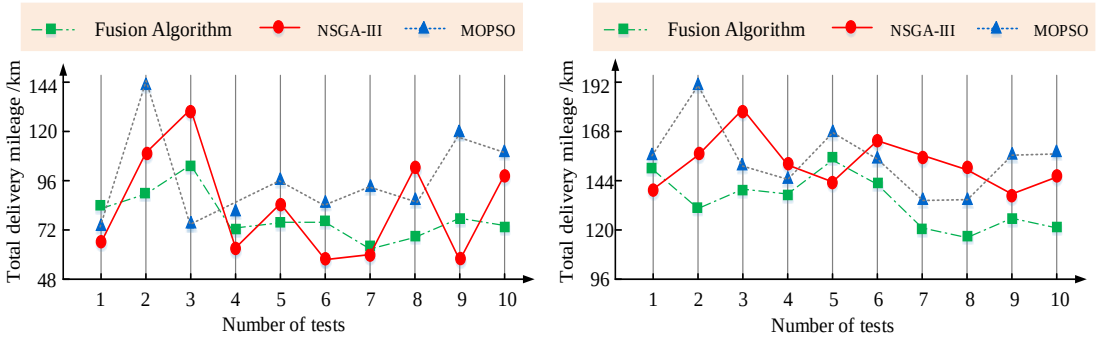


Figure 10: Comparison of distribution process between urban and suburban areas

Figures 10 (a) and (b) show a comparison of 10 delivery distances for urban and suburban customers using different methods. In Figure 10 (a), in the delivery of urban customers, the fusion algorithm shows shorter total delivery mileage in most testing rounds, and its results have higher stability and uniformity in 10 tests. In contrast, NSGA-III and MOPSO show higher peak delivery total mileage in some rounds, especially in the second and third tests, with delivery mileage exceeding 140 km and 130 km, respectively. The fusion algorithm always maintains a low level, with a maximum value of about 100 km. In Figure 10 (b), the fusion algorithm also demonstrates superior performance in the delivery of suburban customers, with a maximum value of about 150 km. Compared with the NSGA-III method and MOPSO method, which have repeatedly exceeded 150 km, the fusion algorithm has significant advantages. In summary, the fusion algorithm has more advantages in delivery in urban and suburban areas, with a shorter total mileage. The distribution path planning results of the fusion algorithm and NSGA-III are shown in Figure 11.

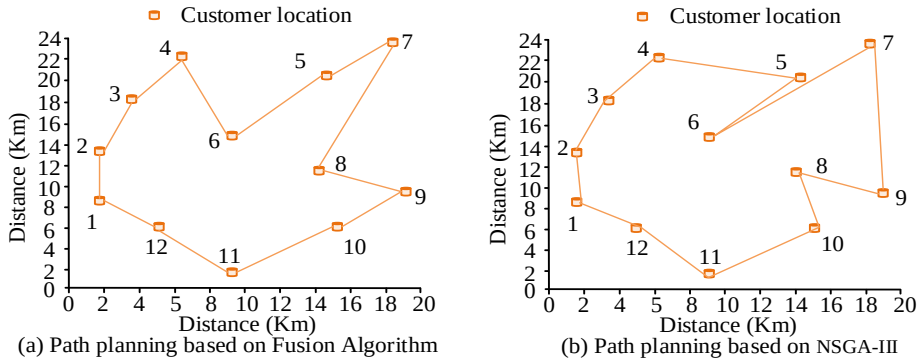


Figure 11: Delivery path planning in two scenarios

In Figure 11 (a), each delivery path planned by the fusion algorithm has no overlapping parts. The path planned by NSGA-III in Figure 11 (b) has overlapping parts. Figure 11 intuitively shows that the delivery route planned by the fusion algorithm has a reduced route length compared to NSGA-III, and avoids the cost loss led by

path duplication. This study compares the fusion algorithm, NSGA-III, and MOPSO methods using average delivery cost and delivery efficiency as indicators, as shown in Figure 12.

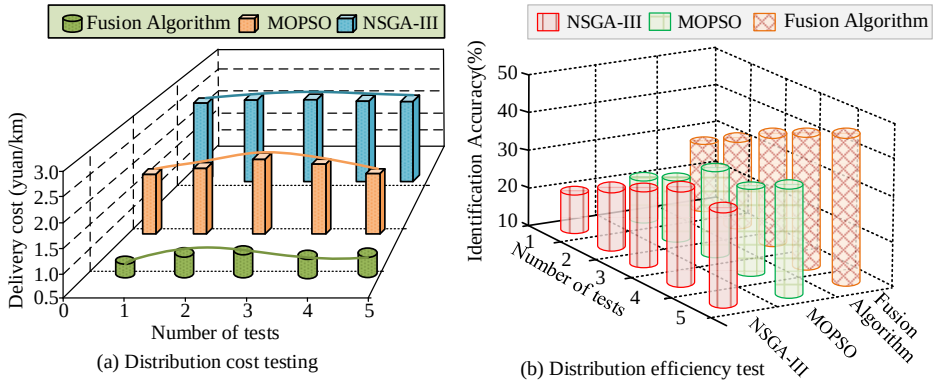


Figure 12: Distribution cost and efficiency test results

Figure 12 (a) shows the test results of the distribution costs of different methods. Figure 12(b) shows the average driving speed of distribution vehicles of different methods, which reflects the distance that vehicles can complete in a unit time. The higher the average speed, the longer the distance that vehicles can cover in the same time, indicating higher distribution efficiency. In Figure 12 (a), the delivery cost of the fusion algorithm remains at the lowest level, with an average cost of about 1.2 yuan per kilometer. The average cost per kilometer for MOPSO is about 1.8 yuan, while for NSGA-III it is about 2.3 yuan. This indicates that fusion algorithms have significant advantages in reducing delivery costs. In Figure 12(b), the NSGA-III method achieves an average speed of approximately 30 km/h, while the MOPSO method reaches about 28 km/h. Notably, the hybrid algorithm demonstrates the most efficient performance with an average speed of around 40 km/h. In summary, fusion algorithms can not only effectively reduce delivery costs, but also achieve significant improvements in delivery efficiency, especially demonstrating excellent adaptability in complex delivery scenarios.

4 Discussion

The proposed fusion algorithm demonstrated significant advantages over NSGA-III and MOPSO in multiple aspects. The results of Table 1 and Figures 9-12 indicated that the algorithm outperformed traditional algorithms in key indicators such as average processing time, resource utilization, and energy consumption. Specifically, on the GORD dataset, it reduced average processing time by up to 1.36 seconds compared to NSGA-III and by up to 2.63 seconds compared to MOPSO. Similarly, on the VRPI dataset, it achieved reductions of up to 0.74 seconds and 2.01 seconds. The notable decrease in resource utilization indicated more efficient resource allocation. In terms of energy consumption, compared with NSGA-III, the fusion algorithm reduced energy consumption by 6.15 j and 3.57 j on the GORD and VRPI datasets, and by 12.8 j and 13.57 j compared to MOPSO. The fusion algorithm also showed clear advantages in delivery cost and efficiency. Figures 10 and 12 demonstrated lower delivery costs and higher efficiency in both urban and suburban deliveries. The average delivery cost per kilometer was significantly lower than NSGA-III and MOPSO, with delivery efficiency notably improved. These results highlighted the practical benefits of the fusion algorithm in real-world logistics operations. Statistical analysis of the study's outcomes confirmed the significance of these performance differences. The fusion algorithm consistently outperformed NSGA-III and MOPSO across multiple iterations and datasets. The reductions in processing time, resource utilization, and energy consumption were not coincidental but demonstrated a more efficient optimization approach. Statistical analysis confirmed that the fusion algorithm achieved significant performance improvements compared to existing methods such as NSGA-III and MOPSO, highlighting its novelty and practical value. When developing a CCL DC location model through practical deployment research, enterprises must consider multiple factors to ensure successful implementation and achieve expected benefits. First, compliance with laws and regulations is essential, requiring data encryption and access control measures to safeguard customer and business data security and privacy. Second, to implement optimized model results, seamless integration with existing warehousing, transportation, and vehicle scheduling systems is necessary. This may involve system interface development, data format conversion, and business process coordination. Finally, enterprises should conduct comprehensive cost-benefit evaluations to balance implementation costs against anticipated benefits, ensuring that optimized logistics operations do not compromise financial stability. Future research can further explore integration with other advanced algorithms or incorporate real-time data to enhance the model's adaptability and robustness in dynamic environments.

5 Conclusion

The prosperity of smart cities has made logistics distribution an important component of urban functions, and its efficiency and sustainability have become a hot

research topic. The traditional logistics model has many shortcomings in terms of efficiency and environmental protection. Therefore, the paper proposed a CCL-DCL model built on the fusion of K-means clustering and GA, aiming to enhance CCL delivery efficiency and cut down delivery costs. The model evaluated the clustering effect through contour coefficients and determined the optimal DCL. Subsequently, it combined the global search capability of GA to construct a multi-objective optimization model that comprehensively considers constraints such as real-time traffic status, cargo preservation time, and vehicle load limitations. In the experiment, the fusion algorithm performed well in optimizing the CCL delivery path. The average task processing time was controlled between 6.32 seconds and 8.42 seconds. The resource utilization rate was the lowest, only 75.21% to 78.46%, while the average energy consumption was 149.67 J to 160.72 J. The delivery cost per kilometer of the fusion algorithm was about 1.2 yuan, and the delivery efficiency per kilometer was about 40 km/h. In summary, the path planning model that integrated K-means and GA exhibited excellent adaptability in complex delivery scenarios, avoiding cost losses caused by route duplication. However, the study has inherent limitations. In real-world operations, demand fluctuations caused by promotional activities and seasonal changes often lead to suboptimal route planning, thereby reducing delivery efficiency. In addition, the real-time response capability of the model is insufficient, and there is a lack of effective data integration mechanism, which makes it difficult to adjust in a timely manner during traffic congestion emergencies, limiting its practicality in dynamic environments and potentially causing delays. In addition, the model has poor adaptability to random events such as weather changes or vehicle failures. This means that when the original plan becomes unfeasible, it cannot quickly generate new routes, thereby reducing the reliability of delivery. Future improvements will include implementing real-time data processing mechanisms, developing dynamic path adjustment algorithms, and building more flexible model frameworks to enhance the robustness and adaptability of the system in complex and constantly changing scenarios.

Nomenclature

Symbol	Description
$d(X,C_i)$	the Euclidean distance calculation in K-means
C_{ij}	the j-th attribute of C_i
C_i	the i-th Cluster Center (CC) point
X_j	the j-th attribute of X
X	a data entity
m	a measure of data objects
$s(i)$	silhouette coefficient of sample i
$a^{(i)}$	average distance from sample i to other samples in the same cluster
$b^{(i)}$	minimum average distance from sample i to samples in neighboring clusters
F_1	the fixed cost calculation consisting of land leasing and personnel salaries
R_j	the unit price of land leasing or the salary of personnel per unit area
Z_j	the decision variable
F_2	the operating cost calculation of CCL DC
$R_j^{(1)}$	the daily operating cost of the logistics hub
X_{ij}	the transportation volume of goods from the supplier to the DC
B_{ij}	the unit transportation cost coefficient between the supplier and the DC.
p	the scale adjustment parameter
F_3	the total logistics cost covering the supply end to the cold chain transit node and then distributed to the end consumption node
C_1	the unit distance transportation cost
B_{jk}	the path complexity coefficient from the DC to the demand point
d_{ij}	the length of the path from the supplier to the transit base
X_{jk}	the transportation volume of materials from the supply end to the logistics node
d_{jk}	the average path length from the transit base to the consumer terminal
F_4	the overall spending of goods damage
C_2	the unit cost of goods damage
θ	the coefficient of cargo damage rate.
v	the transportation speed
F_5	refrigeration cost
$\min F$	objective function
F	the total cost of CCL
$\mathrm{A}\left(\hat{C}_j\right)$	the analysis of the mean traffic convenience characteristics of service nodes within the radiation area
$\hat{C}_j$	the cold chain facility is deployed in the backup location.
$A_{\min}$	the minimum critical value of transportation convenience

	required for the effective operation of the cold chain distribution system
$\hat{t}_{j,l}$	the constraint on CCT time
$\hat{t}_{\max}$	the maximum transportation time
$0,, \aleph(\hat{t}_{j,l} + \hat{t}_{i,j}) < \tilde{T}_{\max}$	the constraint on the spoilage of transported products
$\hat{t}_{i,j}$	the transportation time between the production base and the CCL center
$\tilde{T}_{\max}$	the maximum allowable transportation time limit for cold chain products to maintain freshness
$\sum_{j=0}^{n'} h_{i,j} < C_j', \sum_{l \in L'} \hat{h}_j \ldots \hat{Q}$	the constraint on the supply-demand matching relationship of CCL
L'	the set of CCL centers that can serve demand point j.
$fit(i)$	the fitness of i
μ	a positive real number within the interval $(0,1)$.
$F(i)$	the total cost function value of genetic unit i.
P_x	the probability of being selected
$\tilde{n}$	the population size.
f_x	Individual fitness
$C1$	the offspring chromosomes
$C2$	the offspring chromosomes
$P1$	the parent chromosomes
$P2$	the parent chromosomes
u	the crossover points
s	the crossover points
k	the gene position index

References

[1] Chen Y, Lan H, Wang C, Jia A. An integrated distribution scheduling and route planning of food cold chain with demand surge. Complex & Intelligent Systems, 2023, 9(1): 475-491. https://doi.org/10.1007/s40747-022-00811-9

[2] Turan C, Ozturkoglu Y. A conceptual framework model for an effective cold food chain management in sustainability environment. Journal of Modelling in Management, 2022, 17(4): 1262-1279. https://doi.org/10.1108/JM2-09-2020-0239

[3] Sun J, Karia N B. Innovative approaches to assessing cold chain logistics in B2C E-Commerce environments. Journal of the Knowledge Economy, 2024, 15(3): 14670-14699. https://doi.org/10.1007/s13132-023-01669-z

[4] Tang Y M, Chau K Y, Kuo W T, Liu X X. IoT-based information system on cold-chain logistics service quality (ICCLSQ) management in logistics 4.0. Information Systems Frontiers, 2024, 26(2): 689-708. https://doi.org/10.1007/s10796-023-10393-7

[5] Andoh E A, Yu H. A two-stage decision-support approach for improving sustainable last-mile cold chain logistics operations of COVID-19 vaccines. Annals of operations research, 2023, 328(1): 75-105. https://doi.org/10.1007/s10479-022-04906-x

[6] Bai Q, Yin X, Lim M K, Dong C. Low-carbon VRP for cold chain logist considering real-time traffic conditions in the road network. Industrial management & data systems, 2022, 122(2): 521-543. https://doi.org/10.1108/IMDS-06-2020-0345

[7] Pu X, Lu X, Han G. An improved optimization algorithm for a multi-depot vehicle routing problem considering carbon emissions. Environmental Science and Pollution Research, 2022, 29(36): 54940-54955. https://doi.org/10.1007/s11356-022-19370-0

[8] Yin N. Multiobjective optimization for vehicle routing optimization problem in low-carbon intelligent transportation. IEEE Transactions on Intelligent Transportation Systems, 2022, 24(11): 13161-13170. https://doi.org/10.1109/TITS.2022.3193679

[9] Du J, Wang X, Wu X, Zhou F, Zhou L. Multi-objective optimization for two-echelon joint delivery location routing problem considering carbon emission under online shopping. Transportation Letters, 2023, 15(8): 907-925. https://doi.org/10.1080/19427867.2022.2112857

[10] Li K, Li D, Wu D. Carbon transaction-based location-routing-inventory optimization for cold chain logistics. Alexandria Engineering Journal, 2022, 61(10): 7979-7986. https://doi.org/10.1016/j.aej.2022.01.062

[11] Ashari I F, Nugroho E D, Baraku R, Yanda I N, Liwardana R. Analysis of elbow, silhouette, Davies-Bouldin, Calinski-Harabasz, and rand-index evaluation on k-means algorithm for classifying flood-affected areas in Jakarta. Journal of Applied Informatics and Computing, 2023, 7(1): 95-103. https://doi.org/10.30871/jaic.v7i1.4947

[12] Şenol A. Impkmeans: An improved version of the k-means algorithm, by determining optimum initial centroids, based on multivariate kernel density estimation and kd-tree. Acta Polytechnica Hungarica, 2024, 21(2): 111-131. https://doi.org/10.12700/APH.21.2.2024.2.6

[13] Fu X, Sun Y, Wang H, Li H. Task scheduling of cloud computing based on hybrid particle swarm algorithm and genetic algorithm. Cluster Computing, 2023, 26(5): 2479-2488. https://doi.org/10.1007/s10586-020-03221-z

[14] Choudhury D, Acharjee T. A novel approach to fake news detection in social networks using genetic algorithm applying machine learning classifiers. Multimedia Tools and Applications, 2023, 82(6): 9029-9045. https://doi.org/10.1007/s11042-022-12788-1

[15] Abedpour K, Hosseini Shirvani M, Abedpour E. A genetic-based clustering algorithm for efficient resource allocating of IoT applications in layered fog

heterogeneous platforms. Cluster Computing, 2024, 27(2): 1313-1331. https://doi.org/10.1007/s10586-023-04005-x

[16] Layeb A. ck-means and fck-means: two deterministic initialization procedures for k-means algorithm using a modified crowding distance. Acta Informatica Pragensia, 2023, 12(2): 379-399. https://doi.org/10.18267/j.aip.223

[17] Tirkolaee E B, Torkayesh A E. A cluster-based stratified hybrid decision support model under uncertainty: sustainable healthcare landfill location selection. Applied Intelligence, 2022, 52(12): 13614-13633. https://doi.org/10.1007/s10489-022-03335-4

[18] Wang D. Application of Improved Binary K-means Algorithm in Time and Cost Optimization for Regional Logistics Distribution Center Location. Informatica, 2025, 49(6): 103-116. https://doi.org/10.31449/inf.v49i6.7215

[19] Zhu B, Wang H, Fan M. Constructing small sample datasets with game mixed sampling and improved genetic algorithm. The Journal of Supercomputing, 2024, 80(14): 20891-20922. https://doi.org/10.1007/s11227-024-06263-x

[20] Yang L, Tang Z, Liu S. Research on optimisation method for project site selection based on improved genetic algorithm. International Journal of Industrial and Systems Engineering, 2022, 40(3): 309-324. https://doi.org/10.1504/IJISE.2022.122260