Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Multi-objective genetic algorithm-based molecular screening of exosomes and modeling of their relationship with gene probes

Chuanjie Liang[1,✉], Yangjunjie Wang[2], Tianchu Li[1], Xinxin Xiang[1]

[1] *Center of Translational Medicine, Zibo Central Hospital, Zibo, Shandong, 255000, China*

[2] *Department of Nuclear Medicine and Radiotherapy, Zibo Central Hospital, Zibo, Shandong, 255000, China*

ABSTRACT

In the context of artificial intelligence technology, the current academic research on the relationship between exosome molecular screening and the role of gene probes is relatively weak. Accordingly, this paper formulates a modeling study of exosome molecule screening and its relationship with gene probes under the framework of multi-objective genetic algorithm. The multi-objective genetic algorithm is applied to realize the screening of secretory body molecules, and after the completion of the screening work, the mechanism of the role between exosome molecules and gene probes is investigated by constructing a regression model, and the above theoretical knowledge is applied to empirically analyze the research scheme of this paper. The regression coefficients of exosome molecules and gene probes showed significant correlation at 0.05 level, indicating that the mechanism of action between the two is monotonically increasing, which well reveals the influence of exosome molecules on gene probes.

Keywords: multi-objective genetic algorithm, regression model, exosome molecule, gene probe

1. Introduction

Early diagnosis, early detection and early treatment of disease is the key to enhance the therapeutic effect and reduce the mortality rate of disease. Exosomes are a class of nanoscale vesicles with diameters between 30-150 nm carrying biomolecules such as nucleic acids, proteins, lipids, etc., which are widely distributed in body fluids and carry biological information reflecting the physiological status of their parent cells [1]. Studies have shown that the total amount of exosomes in patients with related diseases such as early stage of cancer is significantly increased and cancer-specific exosomes are present [2-3]. Compared with other biomarkers free in body fluids, the exosome double membrane structure protects the internal biomolecules and enhances the stability of biomarkers [4-5]. Therefore,

✉ Corresponding author.
E-mail address: 18678182636@163.com (C. Liang).
Received 10 January 2024; Revised 15 February 2024; Accepted 25 December 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-384

exosomes are considered as emerging liquid biopsy biomarkers for early diagnosis of diseases [6].

Some scholars have proposed to detect total exosomes and cancer source-specific exosomes in the samples in order to realize the identification of clinical samples from patients with early stage of cancer, and to provide a new reference direction for the early diagnosis of cancer. Khodashenas, S. et al. showed that the cell surface enzyme-linked immunosorbent (ELISA) assay technique is a viable method for the detection of exosomal particles in biological samples, and that the use of this technique plays an important role in assessing the ability of exosomes to be used in diagnostic and therapeutic applications [7]. Iha, K. et al. used ELISA assay to isolate luminal and membrane fractions of exosomes, and then quantitatively extracted trace proteins from them for cancer analysis to broaden the understanding of tumor microenvironment [8]. Yokoyama, S. et al. examined the application of ELISA assay in the molecular assessment of serum exosomes, and based on the optimization of the test buffer formulation, the proposed method achieved accurate results during the extraction of serum exosomal carcinoembryonic antigen [9]. Logozzi, M. et al. described an exosome ELISA method based on immunocapture, whose application in plasma samples can effectively address issues in tumor screening, diagnosis and prognosis [10]. McNicholas, K. and Michael, M. Z. point out that the determination of the size and concentration of extracellular vesicles is an important prerequisite for the development of biomarkers for disease diagnosis and therapeutic vectors, and propose a nanoparticle tracking assay for measurements based on the physical properties of the exosomes and related characterizations [11]. Sharma, V. et al. used a nanoparticle tracking analysis strategy to study the mechanism of salivary neuronal exosomes in neurodegenerative disorders and thus diagnose related neurological dysfunctional diseases [12]. Liu, C. et al. proposed the surface plasmon resonance (SPR) method oriented towards exosomal protein analysis, which provides high-resolution diagnostic results for clinical applications as a highly sensitive, label-free, real-time optical detection method [13]. Sina, A. A. I. et al. similarly pointed out that SPR biosensor real-time and label-free assays can show a high degree of specificity from complex biological samples, and that this high specificity and sensitivity feature plays an important role in clinical practice [14]. However, the low abundance of exosomes and cancer-derived exosomes in the early stages of cancer, and the limited sensitivity and low stability of the above traditional detection techniques limit the application of exosome detection in the clinical diagnosis of cancer. Therefore, it is of great significance to establish a convenient, rapid, sensitive, accurate and reproducible detection method.

Nucleic acid aptamers have the advantages of high specificity, high affinity and good stability, and the introduction of signal-labeled antibodies or aptamers to quantify exosomes can improve the specificity, and the use of gene probes has the advantages of easy and quick operation, low cost and high sensitivity [15-18]. However, the conventional gene probe measurement method is easily affected by external factors, especially the large influence of the sample background, and the relationship model between exosome molecules and gene probes needs to be established to enhance the accuracy of the measurement [19-21].

In this paper, under the support of the theory of multi-objective genetic algorithm, the overall structure and vein of the exosome molecular screening model are determined, and relevant parameters and function expressions are also set to finally realize the mathematical modeling task of the exosome molecular screening model. Through the application analysis of the exosome molecular screening model, the independent variables of this paper are determined, while the gene probe is set as the dependent variable, and a biomedical professional data website is taken as the source of research data. Before constructing the multiple linear regression model, it is necessary to verify the relevance of the set research variables with the help of Pearle's selected correlation coefficient, and after passing the relevance verification, the construction of the multiple linear regression model on the relationship between the two roles begins. Finally, the research data and the model of this paper are synthesized to empirically investigate the relationship between exosome molecules and gene probes.

2. Exosome molecular screening under the action of multi-objective genetic algorithm

2.1. Multi-objective genetic algorithm

The use of genetic algorithms alone in the work of exosome molecule screening has been found to be ineffective in solving this type of problem. In this regard, the study of exosome molecular screening under the action of multi-objective genetic algorithm is proposed, which is able to reduce the generation of local optimums and increase the speed of convergence of the algorithm. The multi-objective genetic algorithm is designed as follows:

2.1.1. Fast non-dominated sorting. In a multi-objective minimization optimization problem, if there exist two solutions A and B, if all the objective function values A are less than B, then A is better than B, i.e. A dominates B [22]. If none of the all other solutions is better than A, then A is said to be non-dominated or Pareto optimal. The process of non-dominated stratification is as follows:

A modified sorted hierarchical method is implemented on population P_i, where n_p denotes the number of all individuals dominating individual p in the feasible solution space, and S_p is the set of all individuals dominated by p in the feasible solution space. The procedure is as follows:

Step1: Initialize n_p and S_p for all individuals in population P_t, such that $n_p = 0, S_p = \varnothing, p = 1, 2, \cdots, N$.

Step2: Perform a non-domination judgment on n_p and S_p for all individuals in population P_t, assuming that p, q is any two individuals in the population, and if p dominates q, then $S_p = S_p \cup \{q\}, n_q = n_q + 1$. If q dominates p, then $S_q = S_q \cup \{p\}, n_p = n_p + 1$.

Step3: Initialize the current stratification number k of the population to $k = 1$.

Step4: Find the $n_p = 0$ individuals in P_t, remove them from P_t and add them to the stratified set F_k, i.e. $F_k = F_k \cup \{p\}$.

Step5: Determine whether F_k is empty, if not, subtract 1 from n_p corresponding to all individuals S_p in F_k, and make $k = k + 1$.

Jump to Step2 and repeat the above operation. And so on until the whole population is stratified. By iterating and comparing and filtering one by one, multiple layers of non-dominated solutions are obtained, and these Pareto optimal solutions thus constitute the Pareto frontier.

Where black solid circles represent the first layer of Pareto solutions and gray solid circles represent the second layer of nondominated solutions. Hollow circles represent dominated individuals, which will not participate in the generation of a new generation of individuals and may face elimination.

2.1.2. Crowding ordering. Density assessment plays an important role in maintaining the diversity of a population during multi-objective evolution [23]. Crowding distance is the property used to assess the density of neighboring solutions of a given solution, and Fig. 1 shows a schematic diagram of crowding degree. The crowding distance of individual i is the sum of the difference between the distances of individuals $i+1$ and $i-1$ neighboring i on each subgoal. The main steps to calculate the crowding degree distance are as follows:

Step1: The crowding degree i_d of each point is set to 0.

Step2: The individuals in the population are sorted in a non-dominated order with different objectives, where the crowding degree of the two individuals located at the boundary is set to infinity, i.e., $o_d = l_d = \infty$.

Step3: Calculate the crowding degree of other individuals in the population, the formula is as Eq:

$$i_d = \sum_{j=1}^{m} (| f_j^{i+1} - f_j^{i-1} |) \tag{1}$$

Where: i_d: denotes the degree of congestion at point i, $f_j^i + 1$: the function value of the j th

objective function at point $i+1$. f_j^{i-1}: The function value of the j th objective function at point $i-1$, m denotes the number of objectives.

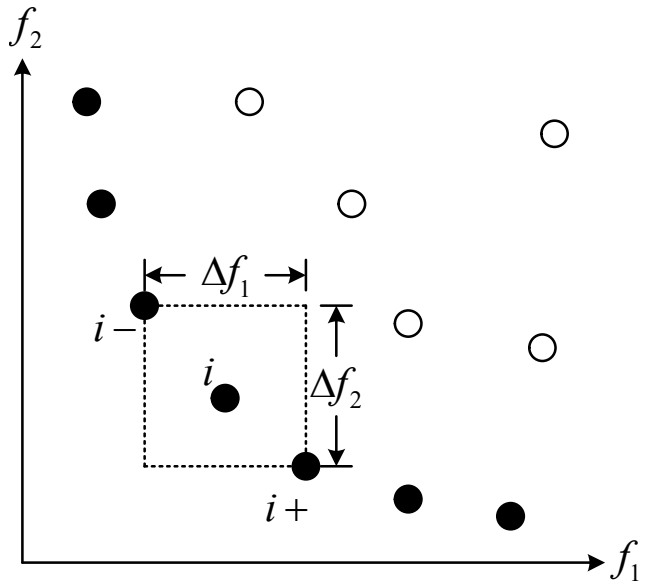


Fig. 1. Examples of congestion calculation

The population will have two attributes, the ordinal number of the non-dominated ordering and the crowding distance, after the fast non-dominated ordering and crowding distance calculations, which are utilized by NSGA-II to compare two individuals i and j in the population and decide which one is better. In general, the solution with a larger non-dominated ordering number or a smaller crowding distance in the NSGA algorithm is more likely to be replaced by other solutions.

2.1.3. Crowding options. After an individual is assigned a non-dominated ordinal number and its crowding degree distance is calculated, in order for the population to gradually approach the Pareto front (PF) during the iteration process, it needs to be accomplished by a selection operation [24]. For Individual A and Individual B, firstly, the individual with small nondominant rank (ordinal number) is retained, and if Individual A and Individual B are in the same nondominant rank, the crowding degree distance between the two is compared, and the individual with smaller crowding degree distance will be eliminated. The mathematical description is:

$$\begin{aligned} &\text{if } A_{rank} < B_{rank}; \\ &\text{else if } A_{rank} = B_{rank} and Ai_d > Bi_d; \\ &A\text{f } B \end{aligned} \tag{2}$$

2.1.4. Elite retention strategies. Elite retention strategies were first proposed for solving single-objective optimization problems. Single-objective problems are limited by population size and random selection, and the excellent solutions are easily lost in the process of selecting the next generation of population. Algorithms using the elite retention strategy are able to retain the excellent solutions generated in each evolutionary generation directly to the next generation, ensuring that the optimal individuals are not replaced in the evolution. Therefore, it improves the convergence efficiency of the algorithm and makes the obtained solutions closer to the Pareto optimal frontier. Usually, a single-objective optimization problem can only obtain one optimal solution, while a multi-objective optimization problem obtains a set of solution sets, so the elite retention strategy for multi-objective problems needs to retain multiple solutions, which is relatively complex.

Among the multi-objective algorithms, the most significant mark of the second generation of multi-

objective evolutionary algorithms is the introduction of the elite retention mechanism. At present, there are two main elite retention strategies: one is the $(\lambda+\mu)$-selection method and the other is the external archiving method. In the $(\lambda+\mu)$-selection method, λ parent populations and μ offspring populations are mixed to form a new population, and the next generation of parent populations is generated by "pruning" method. The principle of "pruning" is to try to keep the best individuals in the mixed population in the next generation with high probability. The external archiving method preserves the Pareto optimal solutions during the evolutionary process by constructing an external archive set. The external archive set is not involved in the optimization operation during the evolution process, but is only used to preserve the optimal solution. During the evolutionary process, the external archive set is iteratively updated according to the Pareto dominance relationship and density information among individuals, subject to the limitation of memory resources and operational efficiency.

In order to improve the convergence efficiency of the NSGA algorithm and retain the excellent individuals in the process of evolution, the NSGA-II algorithm uses the $(\lambda+\mu)$ selection method to retain the elite individuals and selects the offspring individuals by the tournament method. Then the offspring population Q_t and the parent population P_t are merged to form a temporary population R_t, the individuals in the temporary population are non-dominated sorted and the crowding distance of the individuals is calculated, and the excellent individuals from the temporary population are selected as the parent population of the next generation.

2.1.5. Overall process:

1) Generate an initial population, which consists of a certain number of initial individuals, which are called chromosomes.

2) Determine whether the pre-determined end condition is reached, if the condition is reached, output the individual with the best fitness. If the end condition has not been reached, the algorithmic process is entered.

3) Select according to the fitness ordering, the top fitness ordering will be selected to keep, and select some of the lower fitness for subsequent operations.

4) According to the third step to get the group of lower adaptation, according to the crossover rule for crossover.

5) According to the third step to get the lower adaptation group, according to the mutation rule for mutation.

6) Continue to go back to step 3 for continuous evolution until the end condition is reached.

Where the end conditions can be categorized into the following three: first, when the number of iterations reaches a value once. Second, when the proportion of individuals in the population that have reached a certain level of fitness exceeds a certain percentage. Third, when the fitness function value of the optimal individual exceeds a threshold value.

2.2. Exosome molecular screening models

Aiming at the problems of local optimization and slow convergence of traditional genetic algorithm on exosome molecular screening, this subsection uses multi-objective genetic algorithm to construct exosome molecular screening model, so that the results of exosome molecular screening can be better used in gene probes. This subsection first determines the overall idea of exosome molecular screening model, based on which, the mathematical modeling of exosome molecular screening model is completed. The detailed description is as follows:

2.2.1. General idea. Existing molecular screening methods for exosomes are essentially optimized for a single objective (scoring function), such as energy scoring, shape complementation, or chemical complementation, to name a few. However, many real-world problems contain multiple objectives

(which may conflict with each other) or optimization criteria that should be satisfied simultaneously, and a suitable solution to all of them is difficult to be obtained by single-objective optimization. The exosome molecule screening problem also falls into the category of a multiple-objective optimization problem, e.g., the conformation of the secretory exosome molecule obtained by energy scoring optimization often fails to satisfy other criteria (e.g., it is not a good match in terms of shape or it does not satisfy chemical complementation). Therefore, it is necessary to compromise all the objective functions in exosome molecular screening to obtain a more reasonable conformation, which in turn leads to a more rational comparison among different compounds in exosome molecular screening and improves the efficiency of exosome molecular screening.

2.2.2. Mathematical modeling. The multi-objective optimization model for exosome molecular screening can be written as Eq. (3), including a set of n-parameter decision variables, t objective functions, and m constraints. The objective function and constraints are functions of the decision variables. Namely:

$$\begin{aligned} \min \quad & y = f(x) = (f_1(x), f_2(x), ..., f_l(x)) \\ s.t. \quad & g(x) = (g_1(x), g_2(x), ..., g_m(x)) \le 0 \\ where \quad & x = (x_1, x_2, ..., x_n) \in X \\ & y = (y_1, y_2, ..., y_l) \in Y \end{aligned} \tag{3}$$

Decision vectors $x = \{T_x, T_y, T_z, R_x, R_y, R_z, T_{b1}, L, T_{bn}\}^T$, (T_x, T_{yn}, T_z) and (R_x, R_y, R_z) represent the translation and rotation variables, respectively, for the orientation search of the ligand exosome molecule: T_{bI}, Λ, T_{bn} is the angle of rotation of the n rotatable bonds of the ligand exosome molecule in the ligand conformation search. Similarly, the constraint can be expressed as:

$$\begin{cases} \underline{X} \le T_x \le \overline{X} \\ \underline{Y} \le T_y \le \overline{Y} \\ \underline{Z} \le T_z \le \overline{Z} \\ -\pi \le R_x, R_y, R_z, T_{b1}, \Lambda, T_{bn} \le \pi \end{cases} \tag{4}$$

y is the target vector, which in this paper is also known as the scoring function, and the DOCK energy and DOCK contact scores are used as $f_1(x)$ and $f_2(x)$, respectively; of course, more scoring functions can be used in this method, but the only scoring functions available to us right now are the individual DOCK scores. X is the decision space and Y is called the target space.

The scoring functions used in this paper are from DOCK, and the DOCK energy scoring is based on the scoring function of the AMBER force field, which consists of the van der Waals force and the electrostatic potential. The dielectric coefficient $\varepsilon(r) = 4r$ and the van der Waals force is used as Lennard-Jone6-12 function. To wit:

$$E = \sum_{i=1}^{lig} \sum_{j=1}^{rec} \left(\frac{A_{ij}}{r_{ij}^a} - \frac{B_{ij}}{r_{ij}^b} + 332.0 \frac{q_i q_j}{D r_{ij}} \right) \tag{5}$$

DOCK contact score is the sum of heavy atoms in contact between the ligand and the receptor exosome molecule, contact is defined as the distance between the two atoms at a certain threshold (usually $4.5\dot{A}$ is used), and if the distance is too close and collision occurs with each other, a large penalty value is given, even though the DOCK contact score is a very simple shape-complemented scoring function, but it is often very effective.

In this paper, DOCK energy score and DOCK contact score are used as two objective functions in multi-objective optimization, which represent exactly two types of scoring functions, one for energy and one based on shape complementation. Chemical scoring was not considered because the DOCK

chemical score is based on the DOCK energy score, and tests have shown that the chemical score is not as accurate as the Energy score and is linearly correlated with the Energy score.

Each individual needs to be compared to individuals in the archive population by a $\varepsilon-$ dominant method. Is the number of objective functions considered:

$$B_j(f)=\begin{cases}\left\lfloor (f_j-f_j^{\min})/\varepsilon_j \right\rfloor & \text{for } minimizingf_j \\ \left\lceil (f_j-f_j^{\min})/\varepsilon_j \right\rceil & \text{for } maximizingf_j\end{cases} \tag{6}$$

$f_j^{\min}$ is the smallest possible value of the j nd objective function, and ε_j is similar to the ε defined in $\varepsilon-$ Domination, which is a user-defined tolerance difference between two objective values. The identification array divides the entire j objective function objective space into superboxes with ε_j size. The solution P will $\varepsilon-$ dominate the entire ABCDA region, whereas the ordinary definition of domination only allows P to dominate the PECFP region. Comparing each individual of the newly generated offspring and each individual in the archive population, if both dominate the offspring individual c_i, implying that this offspring individual is dominated by the archive individual $\varepsilon-$, then this offspring individual is undesirable. If the identification array Bei of the offspring individual c_i dominates the identification array B_a of any of the archive individuals a, this archive individual is deleted while the offspring individual c_i is accepted. If neither of the above cases is true, it implies that the offspring is a $\varepsilon-$ non-dominated solution of the archive individual, which needs to be discussed in two cases:

Individuals of the i child share the same B vector as an archive individual, i.e. they are within the same box, and they first need to be checked for nondomination. A child individual is retained if it dominates an individual in the archive or if the child individual is not dominated by an individual in the archive but is closer to the B vector than the archive individual. They occupy the same box or have the same B-vector and are non-dominated by each other and are retained due to closer proximity otherwise they are deleted.ii If the offspring does not share the same B-vector with any of the archive individuals, the offspring individual is accepted. In this way, a well-distributed set of solutions will be obtained since each box on the frontier of the set of Pareto solutions can only be perturbed by a single solution, and at the same time the number of final Pareto solutions, i.e., the size of the final archive population, is finite, and there is no need to put an additional restriction on the size of the archive population, which can be adjusted by ε_j to obtain the desired Pareto solutions.

3. Study design for the relationship between secretory molecules and gene probes

Through the exosome molecule screening model above, the exosome molecules that meet the research requirements of this paper can be screened out, and the screened exosome molecules are set as the research independent variables, while the gene probes are used as the research dependent variables, and the sources of the research data are also supplemented to make the logical structure more rigorous. After completing a series of preparatory work, the multiple linear regression model for the study variables was designed. At this point, the above is the entire content of this chapter. The details are shown below:

3.1. Data sources and variable setting

3.1.1. Variable settings. Taking the above exosome molecule screening results based on multi-objective genetic algorithm as the independent variable in the multiple linear regression model, and the gene probe set as the dependent variable, the following is constructed to reflect the mechanism of secretosome molecule's influence on the gene probe by constructing a multiple linear regression analysis.

3.1.2. Data sources. The above has completed the setting of the research variables, the next to determine the source of the research data, this paper research data from a biomedical professional data site, the data in it has a good authority and academic influence, so that the results of this research has a higher persuasive power.

3.2. Regression models

3.2.1. Least Squares. For the dependent variable x_1, there is a corresponding independent variable y_1, and the prediction of the resulting $\hat{y}_i$, after calculating the value of $e_i = y_i - \hat{y}_i$, and then square it, and then calculate each point in the data, and finally add up all the e_i^2 can be calculated to fit a straight line and the actual value of the error between. The so-called best-fitting straight line is the smaller the value of the calculated sum of squares of the residuals, i.e. the smaller the error the better the fit. In order to calculate the minimum sum of squares of the residuals, you can make $Q = \sum_{i=1}^{n} [y_i - (\alpha + \beta x_i)]^2$, the use of calculus for the principle of the extreme value of the α and β partial derivatives, and make its first-order derivatives equal to 0 to solve α and β, that is, the least squares estimation method:

$$\frac{\partial Q}{\partial \alpha} = 2\sum (y_i - \alpha - \beta x_i)(-1) = 0 \tag{7}$$

$$\frac{\partial Q}{\partial \beta} = 2\sum (y_i - \alpha - \beta x_i)(-x_i) = 0 \tag{8}$$

The computational analysis of multiple linear regression is required when there are two or more independent variables. Multiple linear regression analysis addresses the following issues: determining whether there is a correlation between several specific variables, predicting or controlling the value of another variable based on the value of one or more variables and can be informed about its accuracy [25]. Factor analysis, which are significant and which are minor factors for the common effect between several independent variables of a variable.

In multiple linear regression, a dependent variable begins to be influenced by more than one independent variable to co-influence it, so the form of its equation becomes:

$$y = \beta_0 + \beta_1 x_1 + \cdots \beta_n x_n + \grave{o} \tag{9}$$

$$E(y) = \beta_0 + \beta_1 x_1 + \cdots \beta_n x_n \tag{10}$$

Since an observation in multiple linear regression is no longer a scalar but becomes a vector, the observation of the independent variable becomes $(1, x_{11}, \ldots, x_{1n}), (1, x_{21}, \ldots, x_{2p})$, while the observation of the corresponding dependent variable remains unchanged, and so superimposing each row of these observations becomes a vector or matrix. I.e:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} X == \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix} \tag{11}$$

$$\grave{o} = \begin{bmatrix} \grave{o}_1 \\ \grave{o}_2 \\ \vdots \\ \grave{o}_n \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} \tag{12}$$

After determining the multiple regression equation, the coefficient of determination r^2 and the estimated standard error can be used to determine the goodness of fit of the equation. The coefficient

of determination is used to explain how much of the change in Y is explained by the change in the selected independent variable, and if the value of r^2 is too large, the presence of multicollinearity between the variables needs to be considered. The standard error of estimate is used to explain the average deviation between the actual value of Y and the estimated value of $\hat{Y}_l$ when the independent variables are given.

3.2.2. Significance Test. Multiple linear regression equations and univariate linear models are tested in the same way, both can then be tested for the significance of each partial regression coefficient using t and the entire multiple regression model using F. The significance test of the equation is used to test whether the equation is valid or not, if the test result shows that it is not significant, then the equation is not valid. F test: is used to test the significance of all the independent variables $(x_1,...,x_n)$ in the whole for the linearity of y, F the test needs to look at the SignificantF value. P-value and SignificantF value is generally required to be less than 0.05, and the smaller the result, the more significant it is.

After determining the multiple regression equation, it is also necessary to calculate the variance inflation factor (VIF) of each independent variable to determine whether there is a correlation between the independent variables, if there is a high degree of correlation between the independent variables or complete correlation, then there is a multicollinearity between the independent variables, and the effect of each variable on the dependent variable can not be accurately distinguished. Based on the variance inflation factor it is possible to determine whether multicollinearity exists between each independent variable. Determine whether the VIF of each independent variable is greater than 5, if so, this variable needs to be eliminated If the VIF is less than 5, then it can be assumed that the multiple regression equation that has been determined does not suffer from severe multicollinearity.

Residual analysis is used to evaluate the goodness of fit of the linear regression model to the actual data. When testing the multiple linear regression model, the residual plots of the simple linear regression equations between each independent variable and the dependent variable can be analyzed separately, and if the scatters in the residual plots show a relatively random pattern and it is not obvious that there is a certain regularity in the scatters, then it can be assumed that such multiple linear regression model is valid.

4. Empirical analysis of exosome molecules and gene probes

4.1. Empirical analysis of exosome molecular screening models

4.1.1. Algorithm parameters. The source of data in this paper is 50 exosome molecules (X1~X50 are platelets, serum uric acid, etc.) provided by a website, and the method of mean filling column by column is used for all the vacant values in the original data. After replenishing the data of exosomal molecules, the parameters of the multi-objective genetic algorithm mentioned in this paper are set, the population size is set to 20, the number of iterations of the exosomal molecule screening model is 100, the storage capacity of the individual nondominated solution set ND_i is 100, and the storage capacity of the global nondominated solution set ND_R is 400, the probability of updating mode selection is 0.85, the probability of crossover is 0.85 and the probability of mutation is 0.05.

4.1.2. Analysis of results. For some typical continuous multi-objective optimization problems, the theoretical value of Pareto frontier is often easy to obtain, while for discrete multi-objective optimization problems filtered by real data analysis, the theoretical value of Pareto frontier is often unavailable. In recent years, the “dominance rate” has been used by many scholars to measure the computational performance of multi-objective optimization problems for which the theoretical value of the Pareto frontier cannot be obtained, and this index is used in this paper. The dominance rate is shown in Eq. (13):

$$SC_{AB} = \frac{card\left\{B \mid A \prec B^{'}B^{'} \in B\right\}}{card\left\{B\right\}} \tag{13}$$

Where SC_{AB} indicates the percentage of exosome molecule combinations screened by Algorithm B that are dominated by exosome molecule combinations obtained by Algorithm A. Obviously, the larger the value of SC_{AB}, the better the effect of Algorithm A, and let SC_{AB} indicate the domination rate of Algorithm A over exosome molecule combinations screened by other algorithms.

The average domination rate of exosome molecule screening model is analyzed as shown in Table 1, where N denotes the number of iterative training. From this, it can be concluded that the distribution of the domination rate of exosome molecule combinations screened by this paper's algorithm ranges from 0.6 to 0.9, and its overall domination rate is 0.7407, which indicates that this paper's algorithm has an excellent domination effect on the screening of exosome molecules.

Table 1. Average dominance of exosome molecular screening models

N	Dominance rate	N	Dominance rate	N	Dominance rate	N	Dominance rate
1	0.681	26	0.89	51	0.849	76	0.603
2	0.658	27	0.691	52	0.602	77	0.698
3	0.685	28	0.645	53	0.722	78	0.653
4	0.621	29	0.871	54	0.734	79	0.72
5	0.738	30	0.894	55	0.685	80	0.811
6	0.769	31	0.627	56	0.743	81	0.719
7	0.807	32	0.748	57	0.609	82	0.72
8	0.813	33	0.628	58	0.657	83	0.687
9	0.724	34	0.839	59	0.742	84	0.702
10	0.621	35	0.886	60	0.766	85	0.614
11	0.765	36	0.636	61	0.799	86	0.723
12	0.73	37	0.893	62	0.774	87	0.684
13	0.889	38	0.681	63	0.858	88	0.695
14	0.692	39	0.781	64	0.617	89	0.628
15	0.87	40	0.821	65	0.698	90	0.734
16	0.701	41	0.735	66	0.647	91	0.878
17	0.786	42	0.776	67	0.672	92	0.889
18	0.877	43	0.746	68	0.7	93	0.637
19	0.715	44	0.866	69	0.665	94	0.734
20	0.783	45	0.804	70	0.775	95	0.799
21	0.685	46	0.655	71	0.701	96	0.636
22	0.686	47	0.813	72	0.834	97	0.635
23	0.893	48	0.768	73	0.742	98	0.864
24	0.666	49	0.691	74	0.886	99	0.817
25	0.665	50	0.669	75	0.875	100	0.861

The importance is taken as an important index of exosome molecule screening, using multi-objective genetic algorithm to analyze the above 30 screening, and the results of exosome molecule screening are shown in Table 2. If the importance value is greater than 0.5, the secretory body molecules are retained, otherwise rejected for processing, after a series of screening work, the final 50 exosomes only 6 meet the requirements of the study, which are X5, X13, X21, X27, X37, and X50, respectively, laying a theoretical foundation for the development of the subsequent research work.

Table 2. Exosome molecular screening results

Exosome molecules	Value	Results	Exosome molecules	Value	Results
X1	0.259	Delete	X26	0.144	Delete
X2	0.274	Delete	X27	0.615	Retention
X3	0.376	Delete	X28	0.308	Delete
X4	0.028	Delete	X29	0.225	Delete
X5	0.415	Retention	X30	0.168	Delete
X6	0.161	Delete	X31	0.179	Delete
X7	0.482	Delete	X32	0.318	Delete
X8	0.283	Delete	X33	0.444	Delete
X9	0.466	Delete	X34	0.026	Delete
X10	0.098	Delete	X35	0.042	Delete
X11	0.119	Delete	X36	0.335	Delete
X12	0.105	Delete	X37	0.799	Retention
X13	0.592	Retention	X38	0.089	Delete
X14	0.206	Delete	X39	0.027	Delete
X15	0.366	Delete	X40	0.471	Delete
X16	0.133	Delete	X41	0.331	Delete
X17	0.121	Delete	X42	0.022	Delete
X18	0.271	Delete	X43	0.312	Delete
X19	0.121	Delete	X44	0.107	Delete
X20	0.149	Delete	X45	0.048	Delete
X21	0.693	Retention	X46	0.068	Delete
X22	0.303	Delete	X47	0.355	Delete
X23	0.215	Delete	X48	0.409	Delete
X24	0.303	Delete	X49	0.062	Delete
X25	0.305	Delete	X50	0.666	Retention

4.2. Empirical analysis of the two mechanisms of action

Based on the research analysis above, after the screening of exosomal molecules was realized by using multi-objective genetic algorithm, the exosomal molecules (X5, X13, X21, X27, X37, X50) after the screening were taken as the independent variables of the study, while the gene probe (Y) was taken as the dependent variable to construct a multiple linear regression model, which was borrowed to probe the mechanism of the two actions. The detailed analysis process is shown below:

4.2.1. Descriptive statistical analysis of study variables. With the data sources of the research variables clarified, the descriptive statistical analysis of the set research variables was carried out by using SPSS software, and the descriptive statistical analysis of the research variables is shown in Figure 2. As can be seen from the data performance in the figure, the mean and standard deviation of exosome molecules (X5, X13, X21, X27, X37, X50) were 2.42±0.11, 2.78±0.15, 3.02±0.13, 3.27±0.18, 3.72±0.21, and 4.31±0.16, and that of the gene probe (Y) was 4.63± 0.12, which adequately summarizes the data distribution of the seven study variables and provides sturdy data support for the following research work.

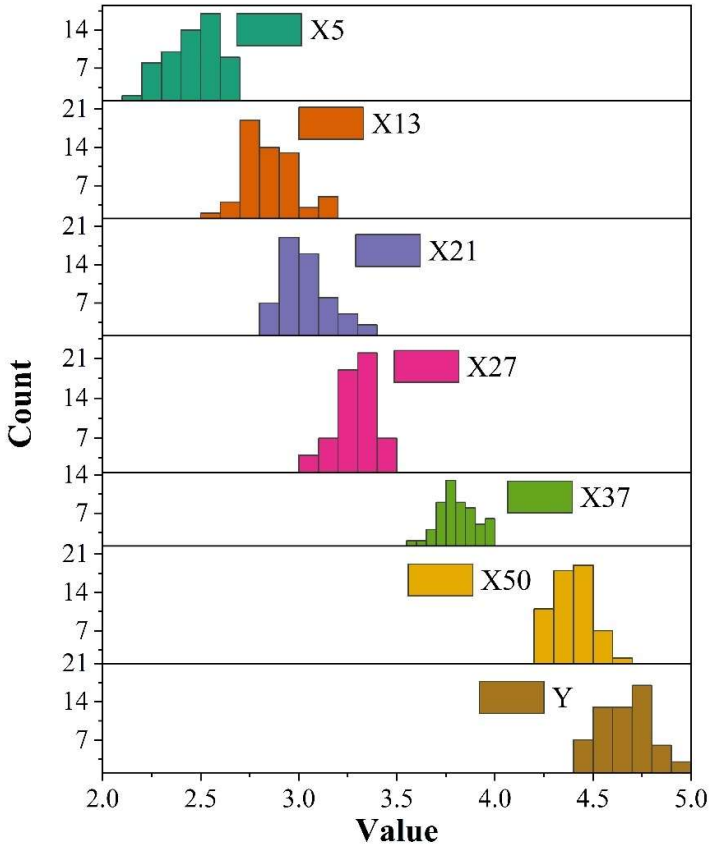


Fig. 2. Descriptive statistical analysis of research variables

4.2.2. Correlation validation analysis. On the basis of the descriptive statistical analysis of the research variables mentioned above, the Pearson correlation coefficient was used to carry out the correlation verification analysis of the set research variables as a way to determine whether the set research variables could be used in the multiple linear regression analysis, and the analysis of the correlation verification results is shown in Fig. 3. Based on the Pearson correlation coefficients in the figure, it can be seen that the Pearson correlation coefficients of exosome molecules (X5, X13, X21, X27, X37, X50) and gene probes were 0.2599 (0.001), 0.1635 (0.006), -0.1023 (0.008), -0.0097 (0.003), -0.0863 (0.017), 0.3274 (0.022), and significant correlation at the 0.05 level, indicating that the research variables set above can be used in the multiple linear regression analysis, and also ensure that the results of the subsequent regression analysis studies are more relevant to the actual situation.

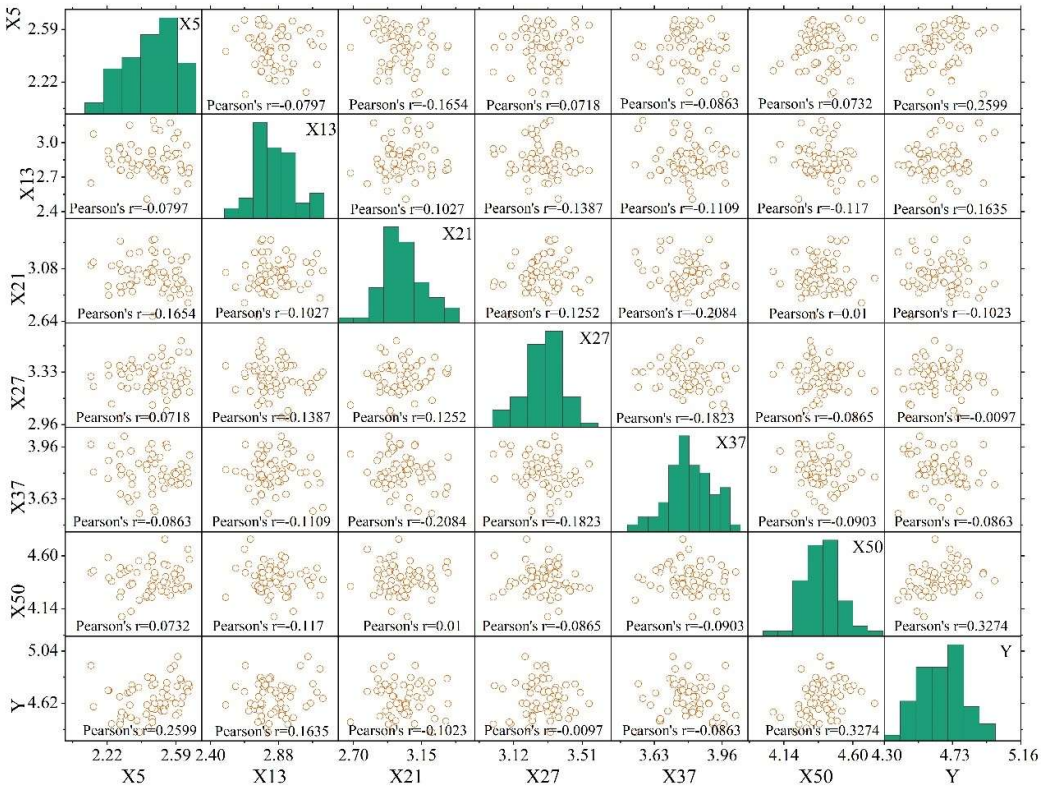
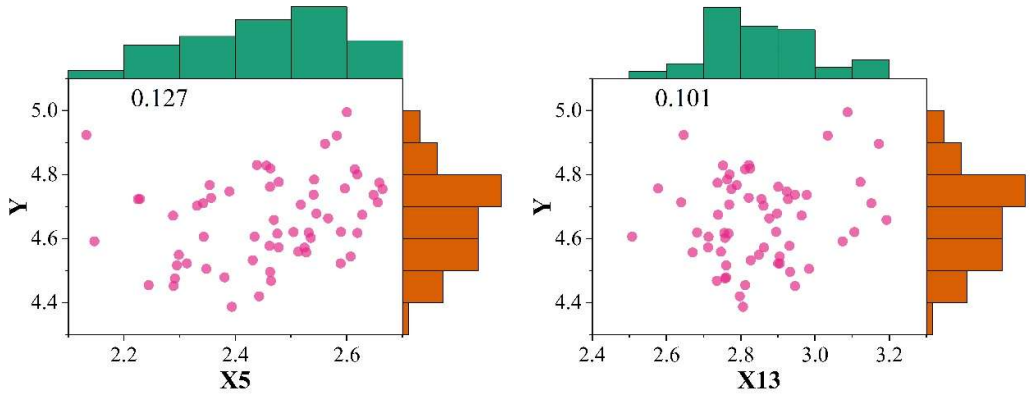


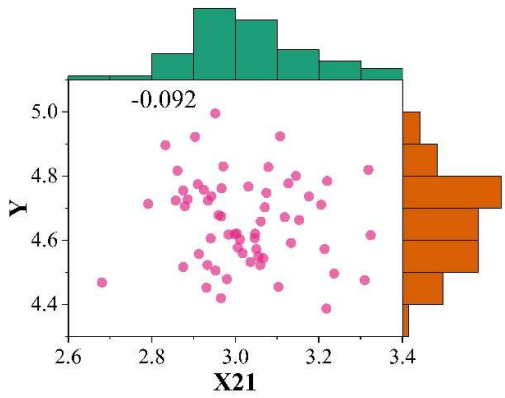
Fig. 3. Analysis of correlation verification results

4.2.3. Multiple linear regression analysis. According to the validation analysis of Pearson correlation coefficient of each research variable above, it can be seen that the independent variables and dependent variables set in the previous article meet the standard requirements of multiple linear regression analysis. At this level, the interaction between exosome molecules and gene probes was analyzed by multiple linear regression using the multiple linear regression model in mathematical statistics theory, and the results of the multiple linear regression analysis are shown in Fig. 4, in which the values represent the regression coefficients of the dependent variables, of which (a) to (f) are X5, X13, X21, X27, X37, and X50, respectively. regression coefficients of variables, it can be concluded that the regression equation of exosome molecules and gene probe is 0.118+0.127*X5+0.101*X13-0.092*X21-0.086*X27-0.072*X37+0.218*X50, and the regression coefficients of the research variables are significant at the 0.05 level, which indicates that exosome molecules have a promotion relationship, which comprehensively reveals the interaction relationship between exosome molecules and gene probes.

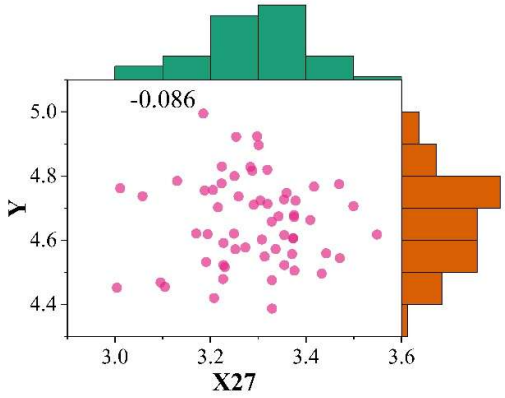


(a)X5

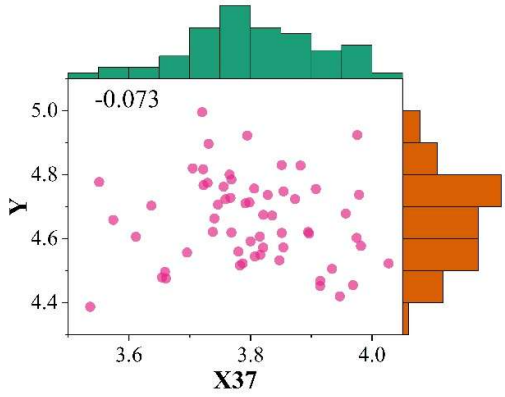
(b)X13



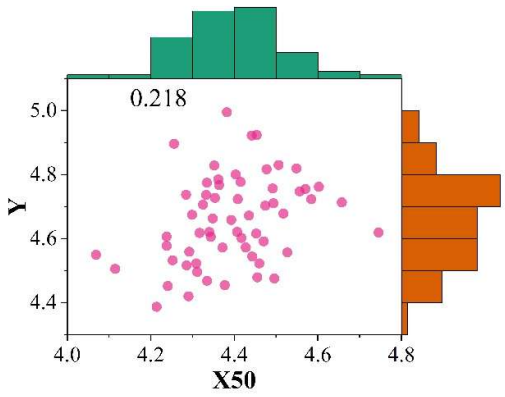
(c)X21



(d)X27



(e)X37



(f)X50

Fig. 4. Multiple linear regression analysis results

5. Conclusion

Aiming at the local convergence phenomenon of traditional genetic algorithm on exosome molecular screening, this paper carries out the research of exosome molecular screening based on multi-objective genetic algorithm. Under the theoretical guidance of multi-objective genetic algorithm, the overall model framework and the idea of vein are determined, and the construction of exosome molecular screening model is completed by combining relevant parameters. The exosome molecular screening results (X5, X13, X21, X27, X37, X50) were set as independent variables, while the gene

probe (Y) was the dependent variable, in addition to determining the data sources of the research variables, and using the multiple linear regression model to evaluate and analyze the influence mechanism between the two. A regression equation of 0.118+0.127*X5+0.101*X13-0.092*X21-0.086*X27-0.072*X37+0.218*X50 was obtained between the two and the regression equation was significantly correlated, which deepened the scholars' knowledge of the relationship between the effects of exosome molecules and gene probes.

Funding

1) The Natural Science Foundation of Shandong Province, China (Grant ZR2024QC222).
2) The Zibo City medical health technology project (No.20240306019).

References

[1] Tai, Y. L., Chen, K. C., Hsieh, J. T., & Shen, T. L. (2018). Exosomes in cancer development and clinical applications. Cancer science, 109(8), 2364-2374.

[2] Soung, Y. H., Ford, S., Zhang, V., & Chung, J. (2017). Exosomes in cancer diagnostics. Cancers, 9(1), 8.

[3] Li, S., Yi, M., Dong, B., Tan, X., Luo, S., & Wu, K. (2021). The role of exosomes in liquid biopsy for cancer diagnosis and prognosis prediction. International journal of cancer, 148(11), 2640-2651.

[4] Cao, Y., Yu, X., Zeng, T., Fu, Z., Zhao, Y., Nie, B., ... & Li, G. (2022). Molecular characterization of exosomes for subtype-based diagnosis of breast cancer. Journal of the American Chemical Society, 144(30), 13475-13486.

[5] Gao, X., Ran, N., Dong, X., Zuo, B., Yang, R., Zhou, Q., ... & Yin, H. (2018). Anchor peptide captures, targets, and loads exosomes of diverse origins for diagnostics and therapy. Science translational medicine, 10(444), eaat0195.

[6] Yu, D., Li, Y., Wang, M., Gu, J., Xu, W., Cai, H., ... & Zhang, X. (2022). Exosomes as a new frontier of cancer liquid biopsy. Molecular cancer, 21(1), 56.

[7] Khodashenas, S., Khalili, S., & Forouzandeh Moghadam, M. (2019). A cell ELISA based method for exosome detection in diagnostic and therapeutic applications. Biotechnology letters, 41, 523-531.

[8] Iha, K., Tsurusawa, N., Tsai, H. Y., Lin, M. W., Sonoda, H., Watabe, S., ... & Ito, E. (2022). Ultrasensitive ELISA detection of proteins in separated lumen and membrane fractions of cancer cell exosomes. Analytical Biochemistry, 654, 114831.

[9] Yokoyama, S., Takeuchi, A., Yamaguchi, S., Mitani, Y., Watanabe, T., Matsuda, K., ... & Yamaue, H. (2017). Clinical implications of carcinoembryonic antigen distribution in serum exosomal fraction—Measurement by ELISA. PLoS One, 12(8), e0183337.

[10] Logozzi, M., Di Raimo, R., Mizzoni, D., & Fais, S. (2020). Immunocapture-based ELISA to characterize and quantify exosomes in both cell culture supernatants and body fluids. Methods in enzymology, 645, 155-180.

[11] McNicholas, K., & Michael, M. Z. (2017). Immuno-characterization of exosomes using nanoparticle tracking analysis. Exosomes and Microvesicles: Methods and Protocols, 35-42.

[12] Sharma, V., Nikolajeff, F., & Kumar, S. (2023). Employing nanoparticle tracking analysis of salivary neuronal exosomes for early detection of neurodegenerative diseases. Translational Neurodegeneration, 12(1), 7.

[13] Liu, C., Zeng, X., An, Z., Yang, Y., Eisenbaum, M., Gu, X., ... & Wu, Y. (2018). Sensitive detection of exosomal proteins via a compact surface plasmon resonance biosensor for cancer diagnosis. ACS sensors, 3(8),

1471-1479.

[14] Sina, A. A. I., Vaidyanathan, R., Wuethrich, A., Carrascosa, L. G., & Trau, M. (2019). Label-free detection of exosomes using a surface plasmon resonance biosensor. Analytical and bioanalytical chemistry, 411(7), 1311-1318.

[15] Mathiyalagan, P., & Sahoo, S. (2017). Exosomes-based gene therapy for microRNA delivery. Cardiac gene therapy: Methods and protocols, 139-152.

[16] Munagala, R., Aqil, F., Jeyabalan, J., Kandimalla, R., Wallen, M., Tyagi, N., ... & Gupta, R. C. (2021). Exosome-mediated delivery of RNA and DNA for gene therapy. Cancer letters, 505, 58-72.

[17] Wan, S., Zhang, L., Wang, S., Liu, Y., Wu, C., Cui, C., ... & Tan, W. (2017). Molecular recognition-based DNA nanoassemblies on the surfaces of nanosized exosomes. Journal of the American Chemical Society, 139(15), 5289-5292.

[18] Liang, Y., Duan, L., Lu, J., & Xia, J. (2021). Engineering exosomes for targeted drug delivery. Theranostics, 11(7), 3183.

[19] Zhang, X., Liu, C., Pei, Y., Song, W., & Zhang, S. (2019). Preparation of a novel Raman probe and its application in the detection of circulating tumor cells and exosomes. ACS applied materials & interfaces, 11(32), 28671-28680.

[20] Agarwal, S., Agarwal, V., Agarwal, M., & Singh, M. (2020). Exosomes: structure, biogenesis, types and application in diagnosis and gene and drug delivery. Current Gene Therapy, 20(3), 195-206.

[21] Han, J., Zhang, L., Cui, M., Su, Y., & He, Y. (2021). Rapid and accurate detection of lymph node metastases enabled through fluorescent silicon nanoparticles-based exosome probes. Analytical chemistry, 93(29), 10122-10131.

[22] Bingzhe Zhang,Kehai Wang,Cancan Yang,Guanya Lu,Yue Li & Haifeng He. (2025). Enhancing seismic performance of highway bridges group with laminated rubber bearings via artificial neural networks and multi-objective genetic algorithm. Structures,74,108470-108470.

[23] Ali Andalibi,Mohammad Safarzadeh,Seyed Mohammad Jafar Sobhani & Hadi Pasdarshahri. (2025). Optimization of a hybrid photovoltaic/thermal-ground source heat pump system using artificial neural networks and multi-objective genetic algorithms. Journal of Building Engineering,103,112103-112103.

[24] Zhijie Yang,Xinkai Hu,Yibing Li,Muxi Liang,Kaipu Wang,Lei Wang... & Shunsheng Guo. (2025). A Q-learning-based improved multi-objective genetic algorithm for solving distributed heterogeneous assembly flexible job shop scheduling problems with transfers. Journal of Manufacturing Systems,79,398-418.

[25] Rui Li,Jianmin Yang,Zilong Cao,Qiang Yue,Hu Jiao,Jia Tian... & Cheng Gan. (2025). A multiple linear regression model calculates the anterior neck skin surface area in Chinese individuals: Implications for neck resurfacing and reconstruction. Journal of Plastic, Reconstructive & Aesthetic Surgery,102,363-372.