Hindawi

Research Article

Survival Prediction Model for Patients with Esophageal Squamous Cell Carcinoma Based on the Parameter-Optimized Deep Belief Network Using the Improved Archimedes Optimization Algorithm

Yanfeng Wang,[1] **Wenhao Zhang** (ID)**,**[1] **Junwei Sun** (ID)**,**[1] **Lidong Wang,**[2] **Xin Song,**[2] **and Xueke Zhao**[2]

[1]*School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450000, China*
[2]*State Key Laboratory of Esophageal Cancer Prevention & Treatment and Henan Key Laboratory for Esophageal Cancer Research of the First Affiliated Hospital, Zhengzhou University, Zhengzhou 450066, China*

Correspondence should be addressed to Junwei Sun; junweisun@yeah.net

Esophageal squamous cell carcinoma (ESCC) is one of the highest incidence and mortality cancers in the world. An effective survival prediction model can improve the quality of patients' survival. Therefore, a parameter-optimized deep belief network based on the improved Archimedes optimization algorithm is proposed in this paper for the survival prediction of patients with ESCC. Firstly, a combination of features significantly associated with the survival of patients is found by the minimum redundancy and maximum relevancy (MRMR) algorithm. Secondly, a DBN network is introduced to make predictions for survival of patients. Aiming at the problem that the deep belief network model is affected by parameters in the construction process, this paper uses the Archimedes optimization algorithm to optimize the learning rate α and batch size β of DBN. In order to overcome the problem that AOA is prone to fall into local optimum and low search accuracy, an improved Archimedes optimization algorithm (IAOA) is proposed. On this basis, a survival prediction model for patients with ESCC is constructed. Finally, accuracy comparison tests are carried out on IAOA-DBN, AOA-DBN, SSA-DBN, PSO-DBN, BES-DBN, IAOA-SVM, and IAOA-BPNN models. The results show that the IAOA-DBN model can effectively predict the five-year survival rate of patients and provide a reference for the clinical judgment of patients with ESCC.

1. Introduction

Cancer is the second leading cause of death in the world and poses a great danger to human health [1, 2]. There will be approximately 19.29 million new cancer cases and 9.95 million cancer deaths worldwide in 2021 [3]. Esophageal cancer is the sixth most common cause of cancer-related death worldwide, including esophageal squamous cell carcinoma and esophageal adenocarcinoma [4]. More than 90% of esophageal cancers are esophageal squamous cell carcinoma (ESCC). The pathology of esophageal squamous cell carci-

noma is complex, and it is often found at an advanced stage, which brings a huge burden to the patient's family [5, 6]. In recent years, the incidence of esophageal squamous cell carcinoma has been increasing [7], and the mortality rate is still high [8, 9].

One of the most fundamental difficulties in the treatment of ESCC is the lack of effective methods for predicting survival risk [10, 11]. Currently, with the more in-depth research on ESCC and the continuous development of medical technology [12], the use of various types of intelligent systems in esophageal cancer diagnosis is increasing [13].

The treatment methods and treatment concepts for patients with ESCC have continued to rise [14]. However, as with other malignancies, the incidence of patients with ESCC is increasing. Even for professional doctors, it is difficult to judge the patient's ultimate risk of survival [15].

Generally, blood indicators, age, and TNM stage information are considered related to the survival rate of cancer patients, and they are often used to predict the survival status of patients [16–18]. In recent years, with the continuous progress of machine learning technology, more and more intelligent algorithms are proposed and applied in multiple fields [19–21]. In the medical field, the research on the survival risk of cancer patients has become a popular research content [22]. A reasonable survival prediction model will effectively improve the survival of cancer patients. Essentially, the cancer patient survival prediction model is a classification problem [23], including the screening of datasets and analyzing the connections between the data. So far, many data mining methods have been proposed in the literature to predict the survival status of esophageal cancer patients [24, 25]. In [26], 90 breast cancer risk miRNAs are predicted based on the proposed DMTN by using the SVM classifier, which obtained an AUV of 0.9633. The method of backpropagation artificial neural network is adopted to predict whether postoperative fatigue occurred in patients undergoing gastrointestinal tumor surgery in [27], and the accuracy rate reached 0.872.

The above approach based on shallow architecture achieves good performance in cancer prediction problems. However, since the classification accuracy of shallow learning depends largely on the quality of the extracted features, it may cause problems when dealing with more complex applications [28]. In fact, for high latitude and complex cancer patient data, it is not sufficient to use simple traditional shallow architecture to solve it [29]. Correspondingly, the deep learning model has multiple nonlinear network structures, which enable it to extract the features of the original data from the hidden layer step by step and improve the classification and prediction accuracy of the model [30, 31]. Therefore, a network structure with deeper layers is preferred.

Deep learning is a new direction in the field of machine learning that models high-level abstractions in input data with hierarchies and multiple layers [32, 33]. Through the establishment of artificial neural network with a network hierarchy, multiple layers gradually extract higher-level features from the original input for learning. Different types of deep neural networks for classification prediction have been used in multiple literatures [34–36]. DBN is a probabilistic generative network, which is considered more suitable for prediction of cancer classification with high feature similarity and complexity [37]. However, in the process of building DBN, improper parameter setting will lead to the instability of the model and the problem of poor classification accuracy. Often, the selection of parameters still relies on the experience of experts to be manually tuned. Aiming at the above problems, a cancer patient survival prediction model based on the improved Archimedes optimization algorithm (IAOA) to optimize DBN parameters is proposed.

In this paper, seventeen blood indicators, age, and TNM staging information of 298 patients with ESCC are studied. Firstly, the clinical data of cancer patients are selected by the minimum redundancy and maximum relevancy algorithm, and the feature indexes are sorted according to their importance. A combination of eleven indicators is selected that is significantly associated with patient survival, which is verified by the Cox regression method in the SPSS software. Secondly, the IAOA is introduced to optimize the parameters in the DBN network training process to improve the stability and classification accuracy of the DBN model. Finally, a survival prediction model of patients with ESCC based on IAOA-DBN is established. The above eleven related indicators are used as inputs, and the five-year survival rate of the patient is used as output. The prediction accuracy rate of IAOA-DBN is better than the existing AOA-DBN, SSA-DBN, PSO-DBN, BES-DBN, IAOA-SVM, and IAOA-BPNN. Therefore, the method for survival diagnosis of patients with ESCC proposed in this paper can accurately predict the survival level of patients. The main contributions of this article can be summarized as follows:

(1) A combination of eleven indicators is found based on minimum redundancy and maximum relevancy feature selection, which is verified to be significantly associated with survival in patients with ESCC

(2) The proposed method uses IAOA to optimize the parameters of the DBN, which effectively improves the stability and classification accuracy of the DBN network. The problem that AOA tends to fall into local optimum and low convergence accuracy is effectively improved by the IAOA. Through the establishment of the IAOA-DBN model, the five-year survival rate of patients with ESCC is effectively predicted

This work is presented as follows. In Section 2, the original data is analyzed, and a combination of multiple indicators that is significantly related to patient survival is found based on minimum redundancy and maximum relevancy algorithm. An improved Archimedes algorithm is proposed in Section 3, which can effectively improve the optimization accuracy and stability of AOA. In Section 4, a survival prediction model based on IAOA-DBN is proposed, which can effectively predict the five-year survival rate of patients with ESCC. In Section 5, the conclusions of this article are presented.

2. Dataset Analysis

2.1. Data Introduction. The clinical data of 298 patients with ESCC used in this article are from patients who were treated in the First Affiliated Hospital of Zhengzhou University from January 2007 to December 2018. The clinical information includes seventeen blood indicators, age, and TNM staging information. The seventeen blood indicators are basophil count (BASO), eosinophil count (EO), fibrinogen (FIB), platelet count (PTL), albumin (ALB), hemoglobin concentration (HGB), white blood cell count (WBC), monocyte count

(MONO), activated partial thromboplastin time (APTT), globulin (GLOB), red blood cell count (RBC), prothrombin time (PT), lymphocyte count (LYMPH), neutrophil count (NEUT), total protein (TP), international normalized ratio (INR), and thrombin time (TT). The population proportion information of the dataset is shown in Table 1. Information of seventeen blood indicators is shown in Table 2.

Among all patients, 147 patients survived more than five years, 151 patients survived less than five years, and the data are evenly distributed. The age distribution of the patients ranged from 38 to 82 years, including 190 male patients and 108 female patients. In addition, the selected patients should have complete treatment records and be followed up for more than six months.

2.2. Minimum Redundancy and Maximum Relevancy Algorithm.
The minimum redundancy and maximum relevancy (MRMR) algorithm [38] is a typical feature selection method. The purpose of MRMR is to select the features with the minimal redundancy and the maximal relevance with the class label. The relevance between features and class labels is represented by mutual information. The mutual information is calculated as Equation (1).

$$I(x\,;y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \, dx dy, \quad (1)$$

where x and y are given two random variables, $p(x, y)$ is the joint probability density function of x and y, $p(x)$ and $p(y)$ are the probability density functions of x and y, respectively. The minimum redundancy and maximum relevancy are calculated as follows, respectively.

$$\max D(S, c)\,; D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i\,; c), \quad (2)$$

$$\min R(S)\,; R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I\left(x_i\,; x_j\right), \quad (3)$$

where S and $|S|$ are feature subsets and the number of features contained therein, respectively, C is the class label, $I(x_i\,; c)$ is the mutual information between feature i and class label C, $I(x_i\,; x_j)$ is the mutual information between feature i and feature j, D is the mean between each feature in the feature set S and the class label C, indicating the relevance between the feature set and the corresponding class label, and R is the size of the mutual information between the features in the feature set S, which represents the redundancy between the features.

The goal of the MRMR algorithm is to maximize the classification performance of the selected feature subset while minimizing the feature dimension. Therefore, it is required that the relevance between the feature subset and the label is the largest, and the redundancy between the features is the least. The minimum redundancy and maximum relevancy are constructed as follows.

$$\max \Phi_1(D, R), \Phi_1 = D - R. \quad (4)$$

TABLE 1: Population proportion information of the dataset.

Project	Category	Number of population	Percentage of population
Genders	Male	186	62.4%
	Female	112	37.6%
Ages	≤61.5	192	64.4%
	61.5	112	37.6%
T stages	T1	42	14.1%
	T2	89	29.9%
	T3	165	55.4%
	T4	2	0.6%
N stages	N0	170	57.1%
	N1	80	26.8%
	N2	34	11.4%
	N3	14	4.7%
TNM stages	1	37	12.4%
	2	139	46.6%
	3	106	35.6%
	4	16	5.4%

The main process of minimum redundancy and maximum relevancy (MRMR) algorithm is as follows.

2.2.1. Step 1: The First Feature Is Selected.
The mutual information between all candidate variables and target variables in the clinical data of esophageal cancer patients is calculated. The feature variable with the largest mutual information is the first feature variable selected.

2.2.2. Step 2: The Second Feature Is Selected.
The redundancy between the selected first feature and the other features is calculated. The feature variable with the least redundancy is the second feature variable.

2.2.3. Step 3: Sequential Selection of Other Features.
Based on the selected two feature variables, the selection of the next feature variable is required to make the selected feature subset have the largest relevance with the target variable and the least redundancy with the selected feature. Therefore, it is necessary to satisfy the minimum redundancy and maximum relevancy criterion of Equation (4). Repeat the calculation of the criteria shown in Equation (4), and add the variables that meet the requirements to the selected feature subset in turn. When the number of selected features meets the requirements, the algorithm ends.

In order to clearly express the MRMR process, the framework of MRMR is shown in Algorithm 1.

2.3. Selection of Optimal Subset Combinations.
The patients' 17 blood indicators, age, and TNM staging information are used as input and five-year survival status as output. The patients' indicators are reordered according to their importance by the MRMR method. The reordered dataset is put into the BP neural network [39], and the classification accuracy of the feature combination is verified by tenfold

TABLE 2: Basic information about seventeen blood indicators.

Variable	Mean	Median (range)	Variance	Standard deviation
BASO	0.050	0 (0-1)	0.014	0.118
EO	0.144	0.1 (0-3)	0.074	0.272
FIB	379.262	362.811 (167.613-909.725)	924.038	30.398
PTL	226.289	227.5 (45-448)	62.902	7.931
ALB	42.077	42 (27-59)	25.055	5.005
HGB	137.742	139 (95-189)	227.216	15.074
WBC	6.564	6.1 (2.18-15.3)	4.077	2.019
MONO	0.406	0 (0-1)	0.091	0.301
APTT	35.929	35.1 (15.4-78.5)	62.479	7.904
GLOB	29.077	29 (17-45)	26.240	5.122
RBC	4.452	4.5 (2.93-6.04)	0.224	0.473
PT	10.322	10.2 (7-16.5)	2.834	1.684
LYMPH	1.930	1.905 (0-8)	0.479	0.692
NEUT	3.864	3.5 (0-10.6)	2.829	1.682
TP	71.070	71 (50-92)	51.971	7.209
INR	0.796	0.78 (0.45-1.64)	0.034	0.185
TT	15.569	15.7 (1.3-46.5)	6.629	2.575

The unit of WBC, LYMPH, GLOB, ALB, RBC, BASO, EO, NEUT, TP, HGB, and PLT is g/L. The unit of PT, TT, APTT is second(s). The unit of FIB is mg/L.

Input: F: original feature set
    $|F|$: the number of features contained in the original feature set F.
    C: class label set
Output: S: feature sorted set
1: $S \longleftarrow \varnothing$;
2: **for** f =1: $|F|$
3: The mutual information of each feature is calculated by Equation(1) and stored in the $Mutual\_set$ set.
4: $f_{max} = \max\_sort\{Mutual_{set}\}$
5: $F \longleftarrow F \setminus f_{max}$; $S \longleftarrow f_{max}$
6: **end for**
7: **for** f =1: $|F|$-1
8: The redundancy between f_{max} and the rest of the features is calculated by Equation(2) and stored in the $Redundancy\_set$ set.
9: $f_{min} = \min\_sort\{Redundancy_{set}\}$
10: $F \longleftarrow F \setminus f_{min}$; $S \longleftarrow f_{min}$
11: **end for**
12: **for** f = 1: $|F|$ −2
13: The minimum redundancy and maximum relevancy of each feature is calculated by Equation(4) and sorted in the $MRMR$ set.
14: $S \longleftarrow S \cup MRMR$
15: **end for**
16: The sorted set S is output.

ALGORITHM 1: Framework of MRMR.

---

cross-validation. When the highest classification accuracy is achieved, the combination with the smallest number of features is the optimal feature combination. The result is shown in Figure 1. When the highest classification accuracy is achieved, the number of features is eleven. Therefore, the features selected in this paper are the first eleven features. The eleven features are TNM stage, BASO, Age, PT, FIB, LYMPH, RBC, TT, PLT, T stage, and GLOB.

2.4. Cox Regression Analysis to Verify the Correlation of Indicators. Cox regression models [40] are widely used in the medical field to analyze the effects of multiple variables on survival status and survival time. In this section, Cox regression models are used to further validate the correlation of selected features with a 5-year survival status and survival time of patients with ESCC. The SPSS 26.0 statistical software is used to make the Cox model. The survival time
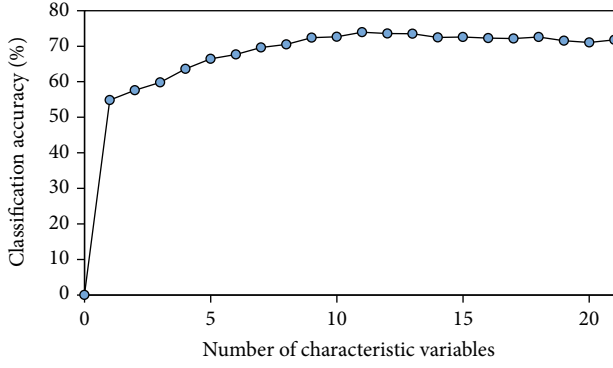
FIGURE 1: The change of classification accuracy with the number of feature variables.

and survival outcome of patients with ESCC are used as dependent variables. The above eleven indicators are independent variables. The survival function at the mean of the covariate is shown in Figure 2. The results show that the p value of the overall score of the eleven indicators is much less than 0.05. The combination of these eleven indicators is significantly related to the survival rate of patients.

3. Improving the Archimedes Optimization Algorithm

3.1. Basic Archimedes Optimization Algorithm. The Archimedes optimization algorithm [41] (AOA) is a new metaheuristic algorithm proposed in 2020. In this algorithm, the population individuals are submerged objects, and the population position is updated by adjusting the density, volume, and acceleration of the objects. According to whether the objects collide in the liquid, AOA is divided into a global exploration stage and a local search stage. If the objects do not collide, the global exploration phase is performed. Instead, a partial development phase is performed.

3.1.1. Initial Stage. In the initialization phase, AOA randomly initializes the density (*den*), volume (*vol*), and acceleration (*acc*) of individuals in the population. The current optimal individual (x_{best}), optimal density (den_{best}), optimal volume (vol_{best}), and optimal acceleration (acc_{best}) are selected. In the AOA, the individual density, volume and transfer factor TF are calculated as Equations (5)–(7), respectively.

$$den_i^{t+1} = den_i^t + \text{rand} \times \left(den_{best} - den_i^t\right), \quad (5)$$

$$vol_i^{t+1} = vol_i^t + \text{rand} \times \left(vol_{best} - vol_i^t\right), \quad (6)$$

where rand is a random number between (0,1). den_i^t and den_i^{t+1} are the densities of the individual i for the generation t and the generation $t+1$, respectively. vol_i^t and vol_i^{t+1} are the volumes of the individual i in the generation t and the generation $t+1$, respectively.

$$TF = \exp\left(\frac{t - t_{\max}}{t_{\max}}\right), \quad (7)$$

where t is the current iteration number and $t_{\max}$ is the maximum iteration number.

When $TF \leq 0.5$, AOA performs a global search, and the update of the individual acceleration is calculated as follows.

$$acc_i^{t+1} = \frac{den_{best} + vol_{best} \times acc_{best}}{den_i^{t+1} \times vol_i^{t+1}}. \quad (8)$$

When $TF = 0.5$, AOA is developed locally, and the individual acceleration is updated to the following:

$$acc_i^{t+1} = \frac{den_{best} + vol_{best} \times acc_{best}}{den_i^{t+1} \times vol_i^{t+1}}. \quad (9)$$

The acceleration of the individual is normalized to obtain Equation (10).

$$acc_{i-\text{norm}}^{t+1} = u \times \frac{acc_i^{t+1} + \min(acc)}{\max(\text{acc}) \times \min(acc)} + l, \quad (10)$$

where $acc_{i-\text{norm}}^{t+1}$ is the normalized acceleration of the individual i in the t generation, u and l are the parameters for adjusting the normalization range.

During the global search phase, the individual positions are updated by Equation (11).

$$x_i^{t+1} = x_i^t + C_1 * \text{rand} * acc_{i-\text{norm}}^{t+1} * d * \left(x_{\text{rand}} - x_i^t\right) \quad (11)$$

where x_i^{t+1} and x_i^t are the positions of individuals in the $t+1$ and t generations and x_{rand} is the positions of random individuals in the generation t. $\text{rand} \in (0, 1)$ is a random number. C_1 is a fixed constant. d is the density factor, which is calculated as follows.

$$d^{t+1} = \exp\left(\frac{t_{\max} - t}{t_{\max}}\right) - \left(\frac{t}{t_{\max}}\right). \quad (12)$$

During the local development stage, the individual position is updated by Equation (13).

$$x_i^{t+1} = x_{\text{best}}^t + F \times c_2 \times \text{rand} \times acc_{i-\text{norm}}^{t+1} \times d \times \left(T \times x_{\text{best}} - x_i^t\right), \quad (13)$$

where c_2 is a fixed constant and F is the direction factor that determines the update direction of the individual position, which is constructed as follows:

$$F = \begin{cases} +1 & \text{if } p \leq 0.5 \\ -1 & \text{if } p\, 0.5, \end{cases} \quad (14)$$

where $p = 2 \times \text{rand} - C_4$ and C_4 is a fixed constant. $T = C_3 \times TF$, and $T \in [C_3 \times 0.3, 1.]$
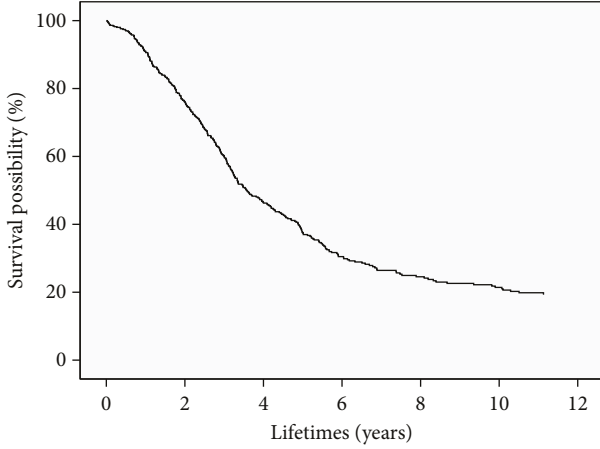
FIGURE 2: Survival function at the mean of the covariate. The survival years are taken as the time, the eleven indicators obtained from minimum redundancy and maximum relevancy algorithm.

3.2. Improved Archimedean Optimization Algorithm.

In the basic AOA, the update of the optimal individual of the population depends on the update of the population in each iteration. After each iteration, the optimal individual is replaced by the individual with the best fitness, and the algorithm does not actively disturb the optimal individual. When the optimal individual of the population falls into the local extremum space, the algorithm will fall into the local optimum, and the phenomenon of premature convergence will occur [42]. Therefore, this paper introduces the corresponding improvement strategy to improve the defects of the basic AOA. Firstly, Sine chaos mapping and reverse learning strategies are used to initialize the population, which can enhance the population diversity and improve the solving efficiency. Secondly, Gaussian variation and superior selection strategies are used to perturb the positions of optimal individuals, which can enhance the global search ability and help the population to jump out of the local optimum. In this paper, the improved AOA is called IAOA. The specific strategy is as follows.

3.2.1. Sine Chaos Reverse Learning Initialization Strategy.

The population of AOA is initialized by random generation. This leads to uneven distribution of individuals in the initial population, which affects the later iterative optimization. The Sine chaotic model [43] is a chaotic model with good randomness and ergodicity with infinite number of map foldings. Reverse learning [44] can obtain its corresponding reverse solution through the current solution. The optimal initial solution can be obtained by comparing and selecting a better solution. In this paper, the Sine chaotic strategy is used to generate an initial population with better diversity. Second, the reverse population is generated according to reverse learning. Finally, the fitness of the obtained population is calculated, and the solution with low fitness is selected as the initial population to improve the probability of obtaining the optimal initial

solution. The 1-dimensional mapping expression of Sine chaos is calculated as the follows.

$$\begin{cases} X_{t+1} = \sin{(2/X_t)} \\ -1 \leq X_t \leq 1, \end{cases} \tag{15}$$

where $t = 0, 1, 2 \cdots, T$ and $X_n \neq 0$.

The population $X = \{X_i, i = 1, 2, \cdots. T\}$, $X_j = \{X_j, j = 1, 2, \cdots \dim\}$ is obtained by mapping the Sine chaos into the solution space. The population individuals are represented as follows.

$$X_{i+1,j} = \sin{(2/X_{i,j})} \tag{16}$$

where $X_{i+1,j}$ is the dimensional j value of the population $i + 1$.

The reverse population can be represented as $X^* = \{X_i^*, i = 1, 2, \cdots T\}$, $X_i^* = \{X_{ij}^*, j = 1, 2, \cdots \dim\}$. The reverse population individual X_{ij}^* can be calculated by the following.

$$X_{ij}^* = X_{\min j} + X_{\max j} - X_{ij}, \tag{17}$$

where $[X_{\min j}, X_{\max j}]$ is the population search dynamic boundary.

The new population $\{X \cup X^*\}$ is formed by the Sine chaotic population X and the reverse population X^*. The fitness values of the new population are ranked, and N individuals with the best fitness values are selected to form the initial population.

3.2.2. Gaussian Operator and Superior Selection Strategy.

The Gaussian operator [45, 46] is introduced in this paper in order to avoid AOA from falling into local optimum and to maintain the diversity of individuals in the population. The current optimal solution X_t^{best} is subjected to Gaussian variation with certain probability p, and a meritocratic selection strategy is taken. The expression of the Gaussian variational operator is calculated as follows:

$$X_i^{t+1} = X_i^t \times (1 + Gauss(\delta)), \tag{18}$$

where X_i^{t+1} denotes the individual position after variation and $Gauss(\delta)$ is a random variable satisfying a Gaussian distribution. The global optimal solution position is updated as follows.

$$X_{\text{best}}^{t+1} = \begin{cases} X_i^{t+1}, \text{otherwise} \\ X_i^t, f(X_i^{t+1}) > f(X_i^t) \text{ and } \mathrm{ran}\,\mathrm{d}_1 < p, \end{cases} \tag{19}$$

where rand_1 is a random variable between $[0, 1]$, p is the probability of superior selection, and $f(.)$ is the individual fitness value. Therefore, variational operations on the global optimal solution can avoid the algorithm from falling into a local optimum and effectively improve the search efficiency.

Input: Initialize algorithm related parameters: the maximum number of iterations $T_{\max}$, population search boundary $[ub, lb]$, parameters C_1, C_2, C_3, C_4, density (den), volume (vol), and acceleration (acc).
1: The population is initialized by using the Sine chaos reverse learning strategy
2: *while* $(t < T_{\max})$
3: *for* **i** = **1** to **N** *do*
4: The den and vol are updated by Equations (5) and (6);
5: The TF is calculated by Equation (7);
6: The d is calculated by Equation (12);
7: if **TF ≤ 0.5 then**
8: The acc_i^{t+1} is updated by Equation (8);
9: The $acc_{i-\text{norm}}^{t+1}$ is updated by Equation (10);
10: The position is updated by Equation (11);
11: *else*
12: The acc_i^{t+1} is updated by Equation (9);
13: The $acc_{i-\text{norm}}^{t+1}$ is updated by Equation (10);
14: The position is updated by Equation (13);
15: *end if*;
16: *end for*
17: The population position is perturbed according to Equation (17)
18: The position is updated when the new position is better than the previous one;
19: $t = t + 1$
20: *end while*
21: The global optimal solution is output

ALGORITHM 2: Framework of IAOA.

In order to clearly express the IAOA process, the framework of IAOA is shown in Algorithm 2.

3.2.3. IAOA Validation and Comparison. In order to fully verify the effectiveness of the IAOA proposed in this paper, the improved Archimedes optimization algorithm, Archimedes optimization algorithm, sparrow search algorithm [47], and bald eagle search algorithm [48] are compared and tested under thirteen benchmark functions at the same time. The selected benchmark functions are classified into three categories. The first category is the single-peak benchmark function, as shown in F1-F5 in Table 3. The second category is the multipeak benchmark function, as shown in F6-F10 in Table 3. The third category is the multimodal benchmark function with fixed dimension, as shown in F11-F13 in Table 3. The basic parameters of the algorithm are as follows: the population size is 30, and the maximum number of iterations is 500. The other parameters within the algorithm are shown in Table 4. The experimental results are presented in Tables 5 and 6. The optimization ability of the algorithm is reflected by the optimal value and the average value, and the stability of the algorithm is reflected by the standard deviation. Firstly, for the five single-peaked functions, IAOA has higher convergence accuracy and stability compared to other algorithms. Secondly, F6 and F8 are able to reach the theoretical optimum when solving for the multipeak function. For other multipeaked functions, IAOA has the best search accuracy and stability. For fixed dimensional functions, IAOA is also better than other algorithms. Therefore, the improvement strategy proposed in this paper has improved the performance of the algorithm to some extent.

TABLE 3: Baseline test functions.

Funs	Name	Range	Optimum
F1	Sphere	$[-100,100]^D$	Min = 0
F2	Schwefel2.22	$[-10, 10]^D$	Min = 0
F3	Schwefel1.20	$[-100,100]^D$	Min = 0
F4	Schwefel2.21	$[-100,100]^D$	Min = 0
F5	Quartic	$[-1.28,1.28]^D$	Min = 0
F6	Rastraign	$[-32, 32]^D$	Min = 0
F7	Ackley	$[-600,600]^D$	Min = 0
F8	Griewank	$[-600,600]^D$	Min = 0
F9	Penalized 1	$[-50, 50]^D$	Min = 0
F10	Penalized 2	$[-50, 50]^D$	Min = 0
F11	Kowalik's	$[-5, 5]^2$	$3.08E-04$
F12	Six-hump	$[-5, 5]^2$	$-1.03E+00$
F13	Branin	$[-5,10]\cup[0, 15]^2$	$3.98E-01$

TABLE 4: Algorithm parameter settings.

Algorithm	Main parameters
AOA	$C_1 = 2$, $C_2 = 6$, $C_3 = 1$, $C_4 = 2$
IAOA	$C_1 = 2$, $C_2 = 6$, $C_3 = 1$, $C_4 = 2$
SSA	ST = 0.6, PD = 0.7, SD = 0.2
BES	$a = 10$; $R = 1.5$

TABLE 5: Comparison with the results of 3 metaheuristic algorithms.

Statistics	Algorithm	F1	F2	F3	F4	F5	F6	F7
Best	IAOA	**$0.00E+00$**	**$0.00E+00$**	**$0.00E+00$**	**$0.00E+00$**	**$5.26E-06$**	**$0.00E+00$**	**$8.88E-16$**
	AOA	$4.93E-126$	$5.61E-60$	$1.27E-100$	$1.55E-59$	$9.09E-05$	$0.00E+00$	$8.88E-16$
	SSA	$0.00E+00$	$0.00E+00$	$0.00E+00$	$0.00E+00$	$3.54E-05$	$0.00E+00$	$8.88E-16$
	BES	$5.31E-46$	$8.70E-29$	$8.10E-19$	$8.13E-14$	$5.36E-05$	$0.00E+00$	$8.88E-16$
Mean	IAOA	**$0.00E+00$**	**$0.00E+00$**	**$0.00E+00$**	**$0.00E+00$**	**$9.89E-05$**	**$0.00E+00$**	**$8.88E-16$**
	AOA	$2.96E-87$	$3.69E-45$	$9.36E-73$	$1.23E-40$	$6.76E-04$	$5.46E+00$	$2.31E-15$
	SSA	$1.58E-64$	$3.17E-29$	$7.63E-42$	$1.62E-45$	$7.50E-04$	$0.00E+00$	$8.88E-16$
	BES	$4.38E-41$	$1.17E-25$	$4.47E-04$	$1.51E-01$	$2.39E-03$	$3.64E+01$	$8.03E-03$
Worst	IAOA	**$0.00E+00$**	**$0.00E+00$**	**$0.00E+00$**	**$0.00E+00$**	**$4.39E-04$**	**$0.00E+00$**	**$8.88E-16$**
	AOA	$5.17E-86$	$1.06E-43$	$2.74E-71$	$2.28E-39$	$2.35E-03$	$1.64E+02$	$4.44E-15$
	SSA	$4.74E-63$	$9.51E-28$	$2.29E-40$	$4.87E-44$	$2.04E-03$	$0.00E+00$	$8.88E-16$
	BES	$4.59E-40$	$1.93E-24$	$1.02E-02$	$6.02E-01$	$8.91E-03$	$1.49E+02$	$2.41E-01$
Std	IAOA	**$0.00E+00$**	**$0.00E+00$**	**$0.00E+00$**	**$0.00E+00$**	**$1.03E-04$**	**$0.00E+00$**	**$0.00E+00$**
	AOA	$8.66E-64$	$1.94E-44$	$5.00E-72$	$4.44E-40$	$5.35E-04$	$2.99E+01$	$1.60E-15$
	SSA	$1.55E-39$	$1.74E-28$	$4.18E-41$	$8.89E-45$	$5.47E-04$	$0.00E+00$	$0.00E+00$
	BES	$1.08E-40$	$3.55E-25$	$1.93E-03$	$2.23E-01$	$2.11E-03$	$5.61E+01$	$4.40E-02$

TABLE 6: Continuation of Table 5.

Statistics	Algorithm	F8	F9	F10	F11	F12	F13
Best	IAOA	**$0.00E+00$**	**$1.04E-13$**	**$4.32E-12$**	**$3.19E-04$**	**$-1.03E+00$**	**$3.98E-01$**
	AOA	$0.00E+00$	$4.85E-01$	$2.60E+00$	$3.35E-04$	$-1.03E+00$	$3.98E-01$
	SSA	$0.00E+00$	$5.02E-06$	$2.21E-05$	$3.08E-04$	$-1.03E+00$	$3.98E-01$
	BES	$0.00E+00$	$7.53E-23$	$8.05E-05$	$3.07E-04$	$-1.03E+00$	$3.98E-01$
Mean	IAOA	**$0.00E+00$**	**$7.03E-09$**	**$6.96E-08$**	**$4.68E-04$**	**$-1.03E+00$**	**$3.98E-01$**
	AOA	$0.00E+00$	$8.16E-01$	$2.89E+00$	$9.10E-04$	$-1.03E+00$	$3.98E-01$
	SSA	$0.00E+00$	$2.42E-05$	$1.49E-03$	$3.74E-04$	$-1.03E+00$	$3.98E-01$
	BES	$0.00E+00$	$1.04E-02$	$5.03E-02$	$3.18E-03$	$-1.03E+00$	$3.98E-01$
Worst	IAOA	**$0.00E+00$**	**$3.86E-08$**	**$3.93E-07$**	**$1.32E-03$**	**$-1.03E+00$**	**$3.98E-01$**
	AOA	$0.00E+00$	$1.15E+00$	$2.99E+00$	$5.91E-03$	$-1.03E+00$	$3.98E-01$
	SSA	$0.00E+00$	$8.30E-05$	$1.32E-02$	$1.22E-03$	$-1.03E+00$	$3.98E-01$
	BES	$0.00E+00$	$1.04E-01$	$2.33E-01$	$2.04E-02$	$-1.03E+00$	$3.98E-01$
Std	IAOA	**$0.00E+00$**	**$1.09E-08$**	**$1.16E-07$**	**$1.68E-04$**	**$1.79E-10$**	**$5.48E-07$**
	AOA	$0.00E+00$	$1.78E-01$	$8.39E-02$	$9.73E-04$	$3.93E-04$	$3.89E-06$
	SSA	$0.00E+00$	$1.87E-05$	$3.88E-03$	$2.53E-04$	$4.72E-08$	$8.16E-07$
	BES	$0.00E+00$	$3.16E-02$	$7.47E-02$	$6.86E-03$	$1.65E-08$	$7.18E-07$

4. Survival Prediction Model of Patients with ESCC Based on IAOA-DBN

4.1. An Overview of DBN. The deep belief network (DBN) is a probabilistic generative network. It is composed by a bunch of restricted Boltzmann machines (RBMs) and a backpropagation (BP) neural network [49]. The learning process of DBN can be divided into pretraining and fine-tuning. During the pretraining process, each RBM is trained individually by an unsupervised learning algorithm in turn, and the network parameters of each layer are gradually adjusted. In the fine-tuning process, the classification labels are used as the output layer of the DBN. The BP neural network is trained sequentially from top to bottom, and the training error is propagated back to the RBM to fine-tune the parameters of all layers to reach the global optimal
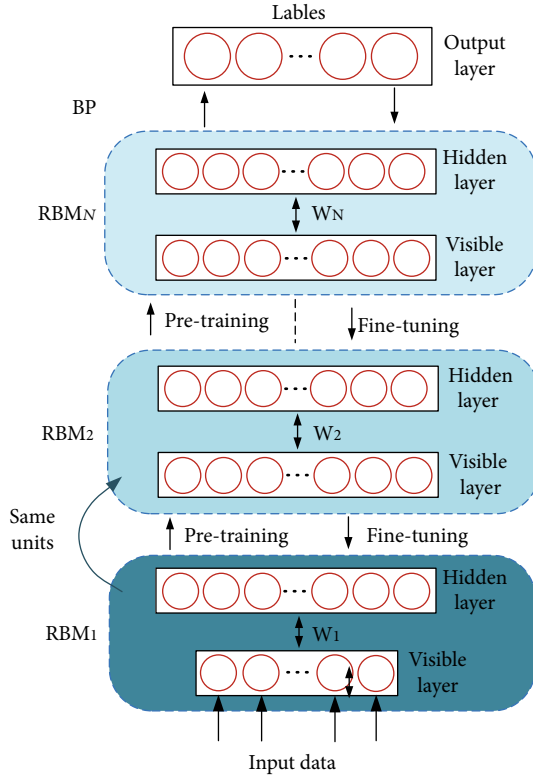
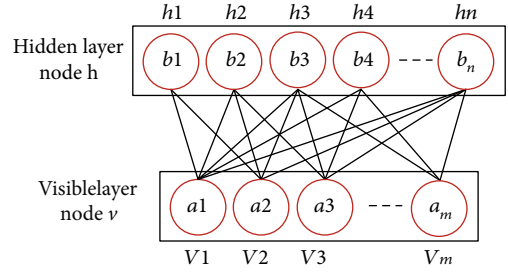FIGURE 3: Schematic diagram of the DBN structure.

FIGURE 4: Schematic diagram of RBM structure.

According to the energy function of the RBM, the joint distribution of the visible layer and the hidden layer is calculated as follows.

$$p(v, h\,;\theta) = \frac{e^{-E(v,h;\theta)}}{T(\theta)}, \tag{21}$$

where $T(\theta) = \sum_{v,h} e^{-E(v,h;\theta)}$, called the normalization factor.

The independent probability distribution of the visible layer is calculated as follows.

$$p(v) = \sum_h p(v, h) = \frac{1}{T(\theta)} \sum_h e^{-E(v,h;\theta)}. \tag{22}$$

There is no connection between nodes in the same layer in the RBM, so the conditional probability distribution of each neuron in the visible layer and the hidden layer is as follows:

$$p\left(h_j = 1|v\,;\theta\right) = \varepsilon\left(\sum_{i=1}^{m} \omega_{ij} v_i + b_j\right),$$

$$p\left(v_i = 1|h\,;\theta\right) = \varepsilon\left(\sum_{j=1}^{n} \omega_{ij} h_j + a_i\right), \tag{23}$$

where $\varepsilon(x) = 1/(1 + \exp(x))$ is the sigmoid function.

The goal of RBM training learning is to make the Gibbs distribution of the RBM network representation as close as possible to the distribution of the original data so that $p(v)$ is maximized.

The network structure parameters $\theta = \{a_i, b_j, \omega_{ij}\}$ of the RBM can be obtained using the maximum likelihood estimation method, and the parameter set θ can be updated by the comparative scattering method, as expressed by the following.

$$L(\theta) = \sum_{t=1}^{T} \log P(v_i|\theta) = \sum_{t=1}^{T} \log \sum_h P(v_i, h|\theta)$$

$$= \sum_{i=1}^{n} \log f(v_i|\theta) - \log T(\theta), \tag{24}$$

parameters of the DBN. The structure of DBN is shown in Figure 3.

4.1.1. Pretraining of RBM. The restricted Boltzmann machine (RBM) is a neural perceptron consisting of a visible layer (v) and a hidden layer (h). Its structure is shown in Figure 4. There are bidirectional connections between the visible and hidden layers, while there is no connection between units in the same layer. In RBM, there is a weight w between any two connected neurons in the visible layer and the hidden layer to represent the connection strength. Each neuron has a bias coefficient a (for the neurons in the visible layer) and b (for the neurons in the hidden layer) to represent its own weight. Therefore, the energy function contained in each RBM is calculated as follows:

$$E(v, h\,;\theta) = -\sum_{i=1}^{m}\sum_{j=1}^{n} \omega_{ij} v_i h_j - \sum_{i=1}^{m} a_i v_i - \sum_{j=1}^{n} b_j h_j, \tag{20}$$

where θ represents the parameter set of RBM, including the state v_i and bias a_i of the visible layer and the state h_j and bias b_j of the hidden layer, ω_{ij} is the connection weight between the visible layer node and the hidden layer node, and n and m represent the number of neurons in the visible layer and the hidden layer, respectively.
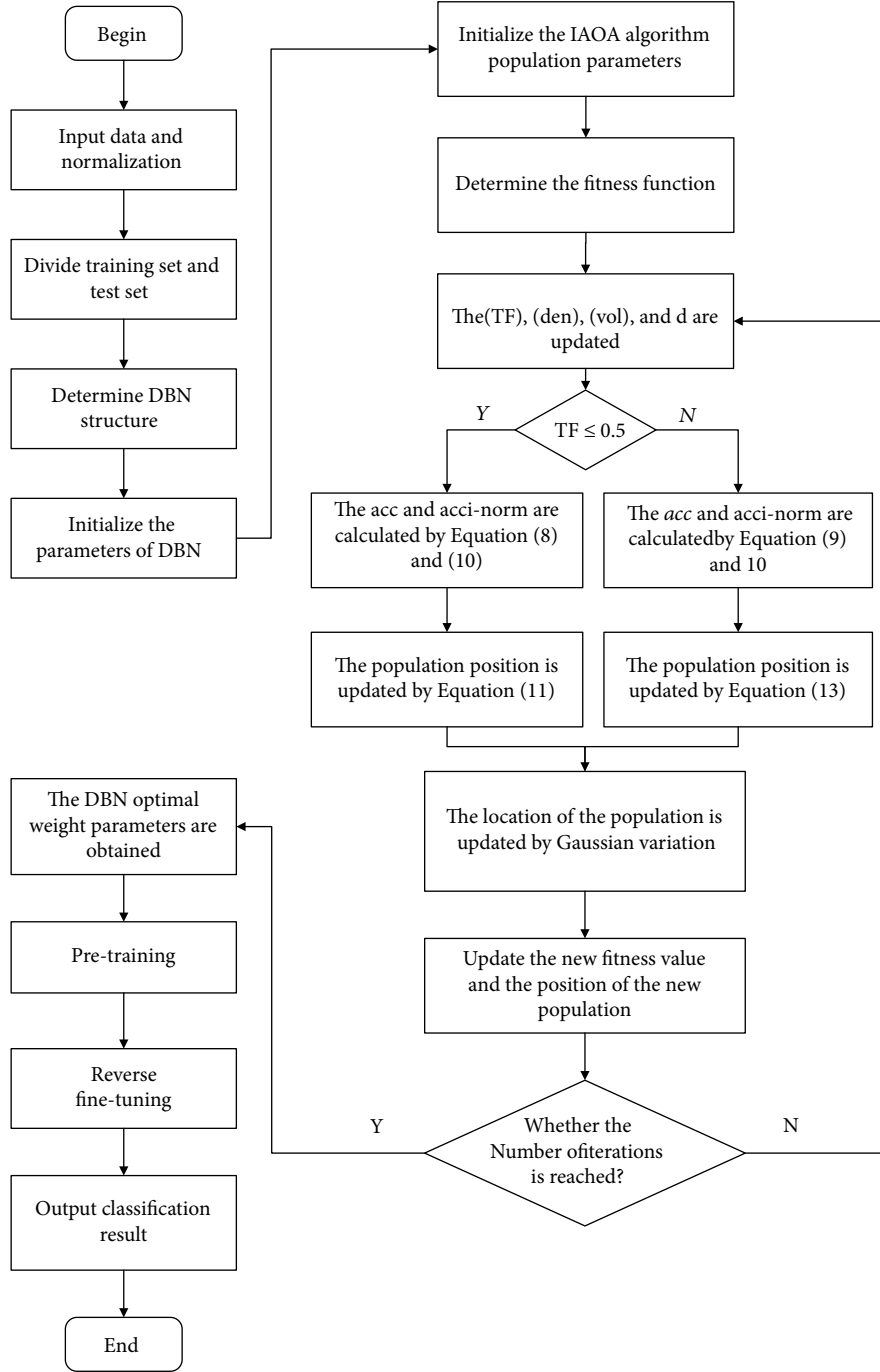
FIGURE 5: The framework of IAOA-DBN.

$$\frac{\partial L(\theta)}{\partial \theta} = \sum_{i=1}^{n} \left[\frac{\partial \log f(v_i \mid \theta)}{\partial \theta} - \left\langle \frac{\partial \log f(v \mid \theta) dv}{\partial \theta} \right\rangle_{p(v \mid \theta)} \right], \tag{25}$$

where $\langle \rangle_{p(v \mid \theta)}$ represents the expected value of the partial derivative under the distribution of $p(v \mid \theta)$.

The model parameter update method is as follows:

$$\omega_{ij}^{t+1} = \omega_{ij}^{t} + \frac{\alpha}{\beta} \left(\langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{remodel}} \right),$$

$$a_i^{t+1} = a_i^{t} + \frac{\alpha}{\beta} \left(\langle v_i \rangle_{\text{data}} - \langle v_i \rangle_{\text{remodel}} \right), \tag{26}$$

$$b_j^{t+1} = b_j^{(t)} + \frac{\alpha}{\beta} \left(\langle h_j \rangle_{\text{data}} - \langle h_j \rangle_{\text{remodel}} \right),$$
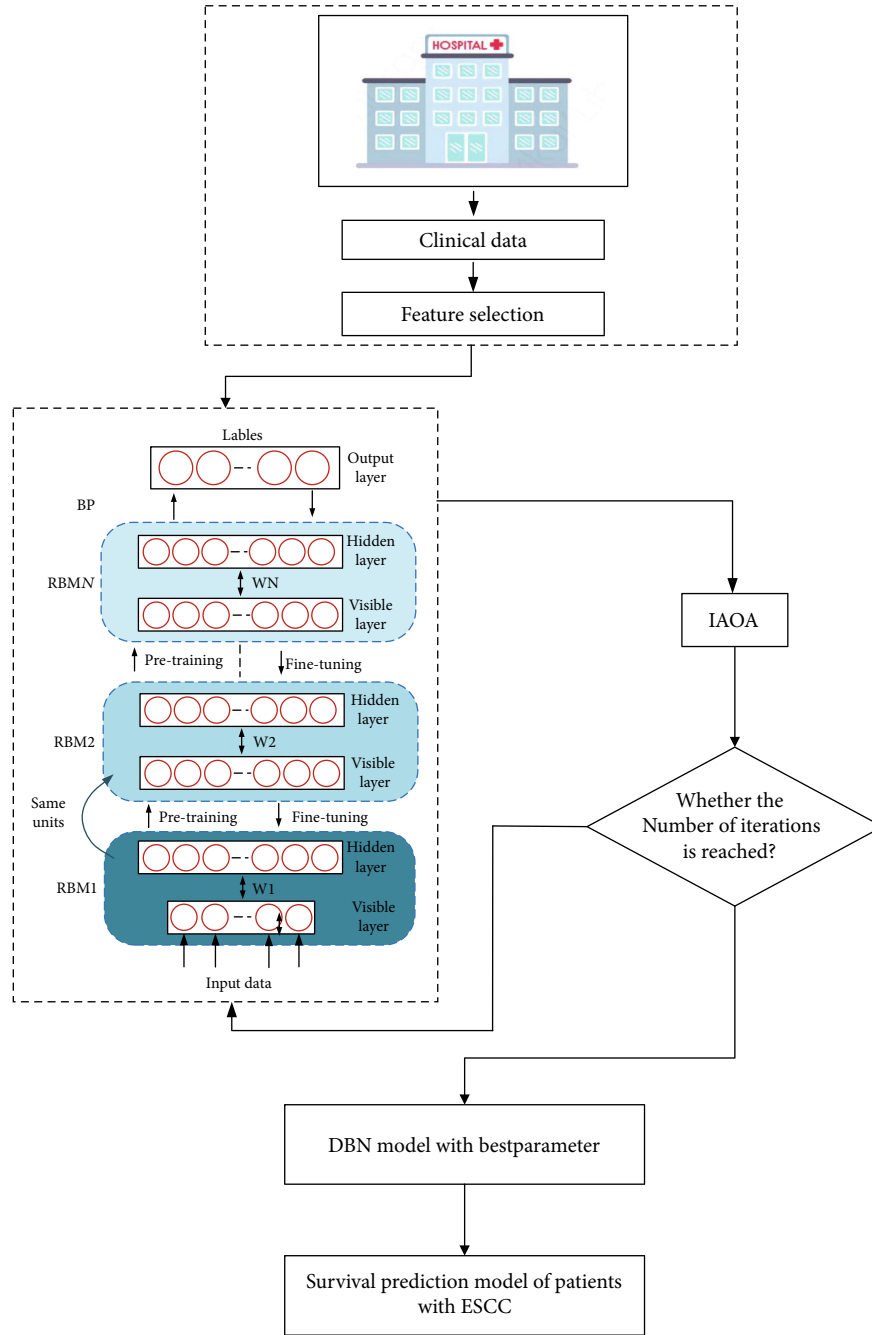
FIGURE 6: ESCC patient survival prediction model.

where $\langle \cdot \rangle_{\mathrm{data}}$ represents the expectation of $p(h\,|\,v)$ defined by the current RBM model, $\langle \cdot \rangle_{\mathrm{remodel}}$ represents the expectation of $p(h\,|\,v)$ defined by the reconstructed RBM model, α represents the learning rate, and β is the batch size.

The pretraining of DBN starts from the bottom layer. After the first RBM is trained, the current hidden layer is transformed into the visible layer of the next RBM. The network is trained layer by layer from bottom to top to avoid falling into local optimum.

4.1.2. Fine-Tuning of RBM. In the fine-tuning stage, a BP neural network is constructed using the hidden layer of the

last RBM and the output layer of the DBN for supervised training. The parameters of each layer are optimized from the top to the bottom to obtain the final model parameters.

4.2. The Proposed Parameter Optimization of DBN Based on IAOA. During the construction of DBN, the choice of hyperparameters such as learning rate α and batch size β has an important impact on the training results of DBN. However, the selection of hyperparameters in traditional DBNs mainly relies on subjective experience, which also causes the problem of insufficient training efficiency. This also leads to a decrease in the classification accuracy and model stability

TABLE 7: Comparison of different algorithms for predicting five-year survival of patients with esophageal squamous cell carcinoma.

Algorithm	10-fold cross-validation accuracy	
	Eleven indicators	All indicators
IAOA-DBN	89.66%	88.13%
AOA-DBN	87.46%	86.24%
SSA-DBN	88.14%	86.93%
PSO-DBN	86.78%	85.46%
BES-DBN	87.29%	86.12%
IAOA-SVM	86.27%	85.19%
IAOA-BPNN	86.61%	85.32%

TABLE 8: Comparison of the results of different algorithms.

Algorithm	10-fold cross-validation accuracy
IAOA-DBN	97.538%
AOA-DBN	96.974%
SSA-DBN	97.177%
PSO-DBN	96.629%
BES-DBN	96.829%
IAOA-SVM	95.602%
IAOA-BPNN	95.038%

of DBNs. In this paper, in order to reduce the influence of human interference factors and improve the classification accuracy of DBN, the IAOA is proposed to optimize the learning rate α and batch size β of DBN. The classification error rate of DBN is used as the objective function of IAOA optimization, and the objective function is $fitness function = 1 - classification error rate$. The larger the fitness value, the higher the classification effect of DBN. In order to clearly express the IAOA-DBN process, the framework of IAOA-DBN is shown in Figure 5.

4.3. Survival Prediction Model of Patients with ESCC. In this paper, eleven indicators significantly related to the survival rate of patients with ESCC are obtained through the MRMR algorithm, and these indicators are TNM stage, BASO, age, PT, FIB, LYMPH, RBC, TT, PLT, T stage, and GLOB, respectively. Eleven indicators and all indicators of the patients are used as inputs to the IAOA-DBN model, respectively, and the five-year survival rate of the patients is used as the output. A survival prediction model for esophageal cancer patients is established. The established survival prediction model for patients with ESCC is shown in Figure 6. To verify the validity of this model, the Archimedes optimization algorithm-deep belief network (AOA-DBN), sparrow search algorithm-deep belief network (SSA-DBN), particle swarm optimization-deep belief network (PSO-DBN) [50], bald eagle search-deep belief network (BES-DBN), improved Archimedean optimization algorithm-support vector machines (IAOA-SVM), and improved Archimedean optimization algorithm-backpropagation neural networks (IAOA-BPNN) are used for comparison. The initial population of AOA, SSA, PSO, and BES is uniformly set to 20, and the maximum number of iterations is 500. The dataset is divided into ten parts, and the tenfold cross-validation method is used to verify the classification accuracy of the model. The prediction results of the DBN optimized by the five optimization algorithms, IAOA-SVM, and IAOA-BPNN model are shown in Table 7.

When eleven patient indicators are used as input, the Tables 5 and 6 show that the prediction results of IAOA-DBN, AOA-DBN, SSA-DBN, PSO-DBN, BES-DBN, IAOA-SVM, and IAOA-BPNN are 89.66%, 87.46%, 88.14%, 86.78%, 87.29%, 86.27%, and 86.61%, respectively. When all patient indicators are used as input, Table 7 shows that the prediction results of IAOA-DBN, AOA-DBN, SSA-DBN, PSO-DBN, BES-DBN, IAOA-SVM, and IAOA-BPNN are 88.13%, 86.24%, 86.93%, 85.46%, 86.12%, 85.19%, and 85.32%, respectively. The comparison shows that IAOA-DBN has a high accuracy rate and can accurately predict the five-year survival rate of ESCC patients. In addition, when the input to the model is eleven indicators, the prediction results are better than using all indicators. Therefore, the MRMR-IAOA-DBN model proposed in this paper can better predict the five-year survival of patients with ESCC.

To better demonstrate the effectiveness of the proposed model, the Wisconsin Diagnostic Breast Cancer (WBCD) dataset is used for testing. In Wisconsin Diagnostic Breast Cancer (WBCD) dataset, 30 indexes of patients are used as input, and the benign and malignant tumors of patients are used as output. The dataset is divided into ten parts, and the tenfold cross-validation method is used to verify the performance of the model. The test results are shown in Table 8. From the test results, it can be seen that IAOA-DBN has higher prediction accuracy than other models. Therefore, the survival prediction model proposed in this paper can effectively predict the prognosis of cancer patients.

5. Conclusions

A novel survival prediction model for patients with ESCC is presented in this paper. Firstly, a minimum redundancy and maximum relevancy algorithm is used to screen out indicators significantly correlated with survival in patients, which is validated by the Cox regression analysis. Secondly, an IAOA-DBN model is proposed. The model uses IAOA to optimize the parameters in the DBN training process, which improves the stability and classification accuracy of the DBN model. Finally, the model is applied to the survival prediction model for patients with ESCC. The results of comparison with four methods verify the validity and superiority of the model. The key conclusions are expressed as follows.

(1) The patients' clinical indicators are ranked by importance using the minimum redundancy and maximum relevancy algorithm, and a new subset of features is selected. The experimental results show that the new feature subset is with better prediction results than the all-feature set

(2) Aiming at the problem that poor convergence accuracy and easy to fall into local optimum of AOA, an improved AOA (IAOA) is proposed in this paper. The experimental results show that the improved strategy proposed in this paper improves the performance of AOA to a certain extent

(3) The learning rate α and batch size β of DBN are optimized using IAOA to obtain the optimal parameters, which improved the classification prediction accuracy and stability of the DBN model. Compared with AOA-DBN, SSA-DBN, PSO-DBN, and BES-DBN, the results verify the effectiveness and superiority of the IAOA-DBN model

Data Availability

The datasets presented in this article are not readily available because the data used in the study are private and confidential data. Requests to access the datasets should be directed to Junwei Sun, junweisun@yeah.net.

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

References

[1] J. Ferlay, M. Colombet, I. Soerjomataram et al., "Cancer statistics for the year 2020: an overview," *International Journal of Cancer*, vol. 149, no. 4, pp. 778–789, 2021.

[2] W. Cao, H. D. Chen, Y. W. Yu, N. Li, and W. Q. Chen, "Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020," *Chinese Medical Journal*, vol. 134, no. 7, pp. 783–791, 2021.

[3] S. R. Knight, C. A. Shaw, R. Pius et al., "Global variation in postoperative mortality and complications after cancer surgery: a multicentre, prospective cohort study in 82 countries," *The Lancet*, vol. 397, no. 10272, pp. 387–397, 2021.

[4] S. Urakawa, T. Makino, M. Yamasaki et al., "Lymph node response to neoadjuvant chemotherapy as an independent prognostic factor in metastatic esophageal cancer," *Annals of Surgery*, vol. 273, no. 6, pp. 1141–1149, 2021.

[5] J. C. Wells, S. Sharma, J. C. Del Paggio et al., "An analysis of contemporary oncology randomized clinical trials from low/middle-income vs high-income countries," *JAMA Oncology*, vol. 7, no. 3, pp. 379–385, 2021.

[6] J. Huang, A. Koulaouzidis, W. Marlicz et al., "Global burden, risk factors, and trends of esophageal cancer: an analysis of cancer registries from 48 countries," *Cancers*, vol. 13, no. 1, p. 141, 2021.

[7] T. Sakamoto, M. Fujiogi, H. Matsui, K. Fushimi, and H. Yasunaga, "Comparing perioperative mortality and morbidity of minimally invasive esophagectomy versus open esophagectomy for esophageal cancer: a nationwide retrospective analysis," *Annals of Surgery*, vol. 274, no. 2, pp. 324–330, 2021.

[8] J. A. Ajani, T. A. D'Amico, D. J. Bentrem et al., "Esophageal and esophagogastric junction cancers, version 2.2019, NCCN Clinical Practice Guidelines in Oncology," *Journal of the National Comprehensive Cancer Network*, vol. 17, no. 7, pp. 855–883, 2019.

[9] L. F. Fransen, G. H. Berkelmans, E. Asti et al., "The effect of postoperative complications after minimally invasive esophagectomy on long-term survival: an international multicenter cohort study," *Annals of Surgery*, vol. 274, no. 6, pp. e1129–e1137, 2021.

[10] F. L. Huang and S. J. Yu, "Esophageal cancer: risk factors, genetic association, and treatment," *Asian Journal of Surgery*, vol. 41, no. 3, pp. 210–215, 2018.

[11] R. J. Kelly, J. A. Ajani, J. Kuzdzal et al., "Adjuvant nivolumab in resected esophageal or gastroesophageal junction cancer," *New England Journal of Medicine*, vol. 384, no. 13, pp. 1191–1203, 2021.

[12] S. H. Lin, B. P. Hobbs, V. Verma et al., "Randomized phase iib trial of proton beam therapy versus intensity-modulated radiation therapy for locally advanced esophageal cancer," *Journal of Clinical Oncology*, vol. 38, no. 14, pp. 1569–1579, 2020.

[13] R. Hashimoto, J. Requa, T. Dao et al., "Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video)," *Gastrointestinal Endoscopy*, vol. 91, no. 6, pp. 1264–1271.e1, 2020.

[14] S. E. Vollenbrock, F. E. Voncken, L. W. Bartels, R. G. Beets-Tan, and A. Bartels-Rutten, "Diffusion-weighted MRI with ADC mapping for response prediction and assessment of oesophageal cancer: a systematic review," *Radiotherapy and Oncology*, vol. 142, pp. 17–26, 2020.

[15] T. B. Nobel, N. Dave, M. Eljalby et al., "Incidence and risk factors for isolated esophageal cancer recurrence to the brain," *The Annals of Thoracic Surgery*, vol. 109, no. 2, pp. 329–336, 2020.

[16] K. Okadome, Y. Baba, T. Yagi et al., "Prognostic nutritional index, tumor-infiltrating lymphocytes, and prognosis in patients with esophageal cancer," *Annals of Surgery*, vol. 271, no. 4, pp. 693–700, 2020.

[17] K. J. Napier, M. Scheerer, and S. Misra, "Esophageal cancer: a review of epidemiology, pathogenesis, staging workup and treatment modalities," *World journal of gastrointestinal oncology*, vol. 6, no. 5, pp. 112–120, 2014.

[18] I. Martincorena, J. C. Fowler, A. Wabik et al., "Somatic mutant clones colonize the human esophagus with age," *Science*, vol. 362, no. 6417, pp. 911–917, 2018.

[19] J. Sun, J. Han, Y. Wang, and P. Liu, "Memristor-based neural network circuit of emotion congruent memory with mental fatigue and emotion inhibition," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 15, no. 3, pp. 606–616, 2021.

[20] J. Sun, J. Han, P. Liu, and Y. Wang, "Memristor-based neural network circuit of Pavlov associative memory with dual mode switching," *AEUinternational Journal of Electronics and Communications*, vol. 129, article 153552, 2021.

[21] J. Sun, G. Han, Z. Zeng, and Y. Wang, "Memristor-based neural network circuit of full-function Pavlov associative memory with time delay and variable learning rate," *IEEE transactions on cybernetics*, vol. 50, no. 7, pp. 2935–2945, 2020.

[22] T. D. Azad, A. A. Chaudhuri, P. Fang et al., "Circulating tumor DNA analysis for detection of minimal residual disease after chemoradiotherapy for localized esophageal cancer," *Gastroenterology*, vol. 158, no. 3, pp. 494–505.e6, 2020.

[23] Y. Jiang, X. Liang, W. Wang et al., "Noninvasive prediction of occult peritoneal metastasis in gastric cancer using deep learning," *JAMA Network Open*, vol. 4, no. 1, pp. e2032269–e2032269, 2021.

[24] M. Doja, I. Kaur, and T. Ahmad, "Current state of the art for survival prediction in cancer using data mining techniques," *Current Bioinformatics*, vol. 15, no. 3, pp. 174–186, 2020.

[25] R. Kaviarasi and R. R. Gandhi, "Accuracy enhanced lung cancer prognosis for improving patient survivability using proposed Gaussian classifier system," *Journal of Medical Systems*, vol. 43, no. 7, pp. 1–9, 2019.

[26] M. Qiu, Q. Fu, C. Jiang, and D. Liu, "Machine learning based network analysis determined clinically relevant miRNAs in breast cancer," *Frontiers in Genetics*, vol. 11, p. 1467, 2020.

[27] X. Y. Xu, J. L. Lu, Q. Xu, H. X. Hua, L. Xu, and L. Chen, "Risk factors and the utility of three different kinds of prediction models for postoperative fatigue after gastrointestinal tumor surgery," *Supportive Care in Cancer*, vol. 29, no. 1, pp. 203–211, 2021.

[28] Y. Xu, Y. Zhou, P. Sekula, and L. Ding, "Machine learning in construction: from shallow to deep learning," *Developments in the Built Environment*, vol. 6, article 100045, 2021.

[29] S. Ronoud and S. Asadi, "An evolutionary deep belief network extreme learning-based for breast cancer diagnosis," *Soft Computing*, vol. 23, no. 24, pp. 13139–13159, 2019.

[30] Y. Xiao, J. Wu, Z. Lin, and X. Zhao, "A deep learning-based multi-model ensemble method for cancer prediction," *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 1–9, 2018.

[31] A. J. de Groof, M. R. Struyvenberg, J. van der Putten et al., "Deep-learning system detects neoplasia in patients with Barrett's esophagus with higher accuracy than endoscopists in a multistep training and validation study with benchmarking," *Gastroenterology*, vol. 158, no. 4, pp. 915–929.e4, 2020.

[32] M. Takeuchi, T. Seto, M. Hashimoto et al., "Performance of a deep learning-based identification system for esophageal cancer from CT images," *Esophagus*, vol. 18, no. 3, pp. 612–620, 2021.

[33] D. Ramachandram and G. W. Taylor, "Deep multimodal learning: a survey on recent advances and trends," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 96–108, 2017.

[34] K. Yasaka, H. Akai, A. Kunimatsu, S. Kiryu, and O. Abe, "Deep learning with convolutional neural network in radiology," *Japanese Journal of Radiology*, vol. 36, no. 4, pp. 257–272, 2018.

[35] M. Lamba, G. Munjal, and Y. Gigras, "A hybrid gene selection model for molecular breast cancer classification using a deep

[36] M. Shanid and A. Anitha, "Adaptive optimisation driven deep belief networks for lung cancer detection and severity level classification," *International Journal of Bio-Inspired Computation*, vol. 18, no. 2, pp. 114–121, 2021.

[37] J. Gai, K. Zhong, X. Du, K. Yan, and J. Shen, "Detection of gear fault severity based on parameter-optimized deep belief network using sparrow search algorithm," *Measurement*, vol. 185, article 110079, 2021.

[38] M. Qasim, Z. Algamal, and H. M. Ali, "A binary qsar model for classifying neuraminidase inhibitors of influenza a viruses (h1n1) using the combined minimum redundancy maximum relevancy criterion with the sparse support vector machine," *SAR and QSAR in Environmental Research*, vol. 29, no. 7, pp. 517–527, 2018.

[39] H. Zhang, C. Chen, R. Gao et al., "Rapid identification of cervical adenocarcinoma and cervical squamous cell carcinoma tissue based on Raman spectroscopy combined with multiple machine learning algorithms," *Photodiagnosis and Photodynamic Therapy*, vol. 33, article 102104, 2021.

[40] H. Li, S. Zhang, J. Guo, and L. Zhang, "Hepatic metastasis in newly diagnosed esophageal cancer: a population-based study," *Frontiers in Oncology*, vol. 11, p. 1151, 2021.

[41] F. A. Hashim, K. Hussain, E. H. Houssein, M. S. Mabrouk, and W. Al-Atabany, "Archimedes optimization algorithm: a new metaheuristic algorithm for solving optimization problems," *Applied Intelligence*, vol. 51, no. 3, pp. 1531–1551, 2021.

[42] A. S. Desuky, S. Hussain, S. Kausar, M. A. Islam, and L. M. El Bakrawy, "EAOA: an enhanced Archimedes optimization algorithm for feature selection in classification," *IEEE Access*, vol. 9, pp. 120795–120814, 2021.

[43] H. Zhu, W. Qi, J. Ge, and Y. Liu, "Analyzing devaney chaos of a sine-cosine compound function system," *International Journal of Bifurcation and Chaos*, vol. 28, no. 14, p. 1850176, 2018.

[44] P. Lin, A. Wang, L. Zhang et al., "An improved cuckoo search with reverse learning and invasive weed operators for suppressing sidelobe level of antenna arrays," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 34, no. 2, article e2829, 2021.

[45] M. Huo, Y. Deng, and H. Duan, "Cauchy-Gaussian pigeon-inspired optimisation for electromagnetic inverse problem," *International Journal of Bio-Inspired Computation*, vol. 17, no. 3, pp. 182–188, 2021.

[46] J. Sun, M. Zang, P. Liu, and Y. Wang, "A secure communication scheme of three-variable chaotic coupling synchronization based on DNA chemical reaction networks," *IEEE Transactions on Signal Processing*, vol. 70, pp. 2362–2373, 2022.

[47] J. Xue and B. Shen, "A novel swarm intelligence optimization approach: sparrow search algorithm," *Systems Science & Control Engineering*, vol. 8, no. 1, pp. 22–34, 2020.

[48] H. Alsattar, A. Zaidan, and B. Zaidan, "Novel metaheuristic bald eagle search optimisation algorithm," *Artificial Intelligence Review*, vol. 53, no. 3, pp. 2237–2264, 2020.

[49] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[50] H. Shao, H. Jiang, X. Zhang, and M. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Measurement Science and Technology*, vol. 26, no. 11, article 115002, 2015.