

Research Article

Relationship between Clinicopathologic Variables in Breast Cancer Overall Survival Using Biogeography-Based Optimization Algorithm

Li-Yeh Chuang ¹, Guang-Yu Chen,² Sin-Hua Moi ², Fu Ou-Yang,^{3,4,5}
Ming-Feng Hou ^{3,4,5,6} and Cheng-Hong Yang ^{2,7}

¹Department of Chemical Engineering & Institute of Biotechnology and Chemical Engineering, I-Shou University, Kaohsiung, Taiwan

²Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

³Division of Breast Surgery, Department of Surgery, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan

⁴Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan

⁵Cancer Center, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan

⁶Institute of Clinical Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan

⁷Ph. D. Program in Biomedical Engineering, Kaohsiung Medical University, Kaohsiung, Taiwan

Correspondence should be addressed to Ming-Feng Hou; mifeho@kmu.edu.tw and Cheng-Hong Yang; chyang@cc.kuas.edu.tw

Received 28 August 2018; Revised 27 February 2019; Accepted 28 February 2019; Published 1 April 2019

Academic Editor: Stephen H. Safe

Copyright © 2019 Li-Yeh Chuang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Breast cancer is the most common cancer among women and is considered a major public health concern worldwide. Biogeography-based optimization (BBO) is a novel metaheuristic algorithm. This study analyzed the relationship between the clinicopathologic variables of breast cancer using Cox proportional hazard (PH) regression on the basis of the BBO algorithm. The dataset is prospectively maintained by the Division of Breast Surgery at Kaohsiung Medical University Hospital. A total of 1896 patients with breast cancer were included and tracked from 2005 to 2017. Fifteen general breast cancer clinicopathologic variables were collected. We used the BBO algorithm to select the clinicopathologic variables that could potentially contribute to predicting breast cancer prognosis. Subsequently, Cox PH regression analysis was used to demonstrate the association between overall survival and the selected clinicopathologic variables. C-statistics were used to test predictive accuracy and the concordance of various survival models. The BBO-selected clinicopathologic variables model obtained the highest C-statistic value (80%) for predicting the overall survival of patients with breast cancer. The selected clinicopathologic variables included tumor size (hazard ratio [HR] 2.372, $p = 0.006$), lymph node metastasis (HR 1.301, $p = 0.038$), lymphovascular invasion (HR 1.606, $p = 0.096$), perineural invasion (HR 1.546, $p = 0.168$), dermal invasion (HR 1.548, $p = 0.028$), total mastectomy (HR 1.633, $p = 0.092$), without hormone therapy (HR 2.178, $p = 0.003$), and without chemotherapy (HR 1.234, $p = 0.491$). This number was the minimum number of discriminators required for optimal discrimination in the breast cancer overall survival model with acceptable prediction ability. Therefore, on the basis of the clinicopathologic variables, the survival prediction model in this study could contribute to breast cancer follow-up and management.

1. Introduction

Breast cancer is the most common cancer among women worldwide, with an estimated 1.67 million newly diagnosed cases each year, ranking second for cancer incidence rate and fifth for cause of death from cancer [1]. Worldwide increase in breast cancer incidence represents a sizeable burden on public health services [2]. Consequently, the diagnosis, treatment,

and prognosis of breast cancer has become a vital research concern [3]. Previous studies have reported various prognosis factors for susceptibility [4–9] and the overall survival of patients in breast cancer [3, 10–12].

The clinicopathologic factors for breast cancer prognosis and overall survival include both tumor burden and tumor molecular biological factors. Tumor burden factors are usually defined as tumor size, lymph node invasion,

and lymph vascular and dermis invasion, and tumor molecular biological factors usually include tests for hormonal status—including estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2)—combined with fluorescence in situ hybridization and immunochemistry test results. The Nottingham prognostic index (NPI) was developed to predict prognosis outcomes for various tumor burden situations [13, 14], and the breast cancer severity score was developed for the same purpose but depends mainly on both tumor burden and tumor molecular biological factors [15]. Age at diagnosis is also a factor that is associated with breast cancer prognosis outcome and overall survival [16]. The effect of age on breast cancer progression and mortality might be affected by other clinicopathologic factors [17, 18]. Although various treatments for breast cancer can result in favorable prognosis and survival rates for breast cancer patients [19–23], the radiotherapy, chemotherapy, hormonal therapy, and targeted therapy are highly reliant on clinicopathologic factors [10].

Formerly, analysis of survival benefit or long-term follow-up of breast cancer was based mainly on the common statistical analysis strategy. Along with the improvements in data volume and complex disease causality, machine learning and optimization algorithms have emerged as novel strategies for analyzing breast cancer survival [24–27]. Both statistical and machine learning and optimization algorithm approaches provide theoretical and acceptable explanations for the association between clinicopathologic factors and breast cancer survival benefit. The combined and hybrid use of both statistical and machine learning approaches is a trend in modern biological research.

Biogeography-based optimization (BBO) is a metaheuristic algorithm that was proposed by Simon in 2008 to solve global optimization problems [28]. BBO is an evolutionary algorithm that was inspired by the migration of species between habitats. Specifically, it was inspired by biogeography, which describes (i) the speciation and migration of species between isolated habitats and (ii) the extinction of species [29]. In recent years, the BBO algorithm has been widely used in a myriad of fields, such as to solve the engineering optimization problem. Various BBOs have been proposed to enhance the BBO search ability in specific problems, including blended BBO [30], localized BBO [31], and ecogeography-based optimization [32]. The BBO algorithm is a powerful search technique because it contains both exploration and exploitation strategies based on migration [33].

To explore the clinicopathologic variables of the overall survival of patients with breast cancer, this study analyzed the relationship between the clinicopathologic variables of breast cancer by using a BBO algorithm. We sought to (i) assess the relationship between the clinicopathologic characteristics of overall survival among patients with breast cancer and (ii) to demonstrate the optimization of the overall survival prediction model on the basis of the clinicopathologic variables of breast cancer.

2. Material and Methods

2.1. Data Source and Patients. All data were collected from the single-center Taiwan Breast Cancer Consortium (TBCC)

database, which is prospectively maintained by the Division of Breast Surgery at Kaohsiung Medical University Hospital, Taiwan. Patients who were diagnosed with ductal carcinoma in situ were excluded. In total, 1,896 patients with breast cancer were included and tracked from 2005 to 2017. The prognosis variables in this dataset included age at diagnosis, grade, tumor size (American Joint Committee on Cancer [AJCC] stage), estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), lymph node (AJCC stage), lymph vascular invasion (LVI) status, dermal invasion, perineural invasion, surgical method, radiotherapy, chemotherapy, and hormone therapy. The overall survival term of all participants with breast cancer was tracked from the date of first diagnosis until participant death or study conclusion. The proposed analysis procedure for TBCC dataset is summarized in Figure 1.

2.2. Biogeography-Based Optimization. The BBO algorithm is a population-based optimization algorithm inspired by the natural biogeographical distribution of species [28], which simulates biogeographical species distribution in accordance with the insular migration of species. In biogeography, an area's quality is evaluated by considering suitability index variables (SIVs), including climate, temperature, and humidity. The habitat suitability index (HSI) represents insular quality. A high HSI value indicates that an area is a superior habitat, whereas a low HSI value indicates that an area is an unsuitable habitat. Higher HSI value areas are usually saturated, which means that species encounter difficulty in migrating to these areas and species currently living in these areas are likely to migrate to other areas. By contrast, low HSI value areas are likely to acquire many migrant species come to here. In the BBO algorithm, an unsatisfactory solution results in an area with a low HSI value, whereas a satisfactory solution results in an area with a high HSI value. According to the BBO mechanism, satisfactory solutions are likely to share their SIVs with other solutions and are unlikely to accept SIVs from other solutions. The BBO migration model depicts the migration of species in a habitat. I is the maximum immigration rate (λ), E is the maximum emigration rate (μ), S_{\max} is the maximum number of species that an island can host, and S_0 is the number of species that causes the immigration rate to equal the emigration rate. The immigration rate increases as the solution quality decreases, whereas the emigration rate decreases with species quantity. This linear model describes the way for simulating species migration. The BBO algorithm comprises two primary stages: migration and mutation.

2.2.1. Migration. In migration, the probabilistic model can be used to represent the concepts of emigration and immigration. Consider the probability P_s that a habitat contains S species; P_s changes from time t to time $t + \Delta t$, which can be formulated as

$$P_s(t + \Delta t) = P_s(t) (1 - \lambda_s \Delta t - \mu_s \Delta t) + P_{s-1}(t) \lambda_{s-1} \Delta t + P_{s+1}(t) \mu_{s+1} \Delta t \quad (1)$$

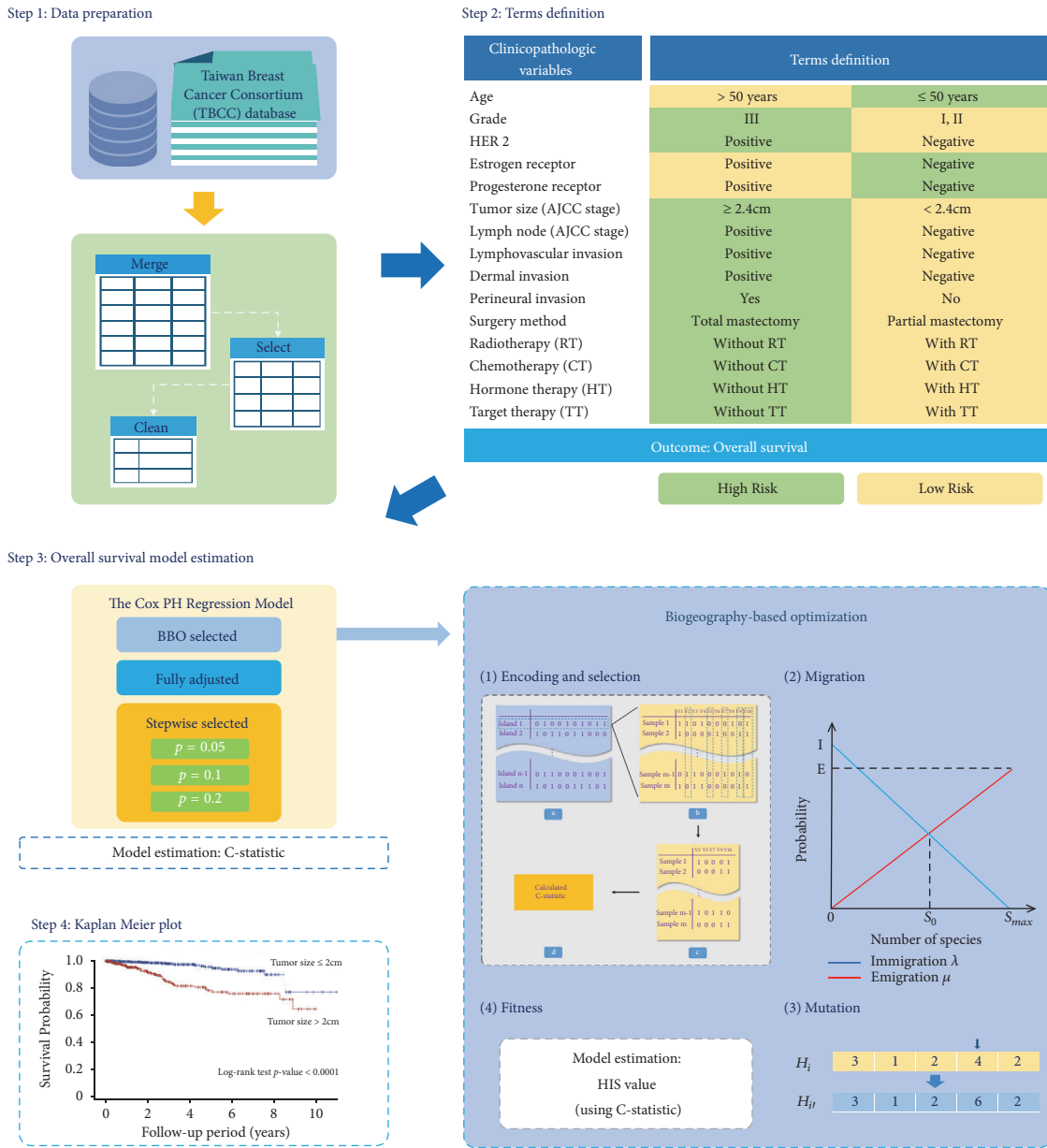


FIGURE 1: Summary of the proposed analysis approaches for Taiwan Breast Cancer Consortium (TBCC) database.

where λ_S and μ_S are, respectively, the immigration and emigration rates when S species are present in the habitat. Equation (1) holds because, to have S species at time $(t + \Delta t)$, one of the following conditions must be true [28]:

- (1) S species were present at time t , and no immigration or emigration occurred between t and $t + \Delta t$.
- (2) $S - 1$ species were present at time t , and one species immigrated.
- (3) $S + 1$ species were present at time t , and one species emigrated.

Suppose that time Δt is sufficiently small; the probability of more than one immigration or emigration occurring can be ignored. Subsequently, taking the limit of (1) as $\Delta t \rightarrow 0$ provides the following equation:

$$\dot{P} = \begin{cases} -(\lambda_S + \mu_S) P_S + \mu_{S+1} P_{S+1}, & S = 0 \\ -(\lambda_S + \mu_S) P_S + \lambda_{S-1} P_{S-1} + \mu_{S+1} P_{S+1}, & 1 \leq S \leq S_{max} - 1 \\ -(\lambda_S + \mu_S) P_S + \lambda_{S-1} P_{S-1}, & S = S_{max} \end{cases} \quad (2)$$

```

1: For each habitat
2:   For each SIV
3:   Select habitat  $H_i$  with probability  $\lambda_i$ 
4:   If  $U(0,1) < \lambda_i$  then
5:     Select  $H_j$  with probability  $\mu_j$ 
6:     If  $U(0,1) < \mu_j$  then
7:       Replace  $H_i(SIV)$  by  $H_j(SIV)$ 
8:     End
9:   End
10: End
11: End

```

ALGORITHM 1: Pseudo code for migration of BBO.

From the migration operation curves, where k number of species are present, the emigration μ_k and immigration λ_k rates can be formulated as (3) and (4), respectively:

$$\mu_k = E \frac{S_k}{S_{max}} \quad (3)$$

$$\lambda_k = I \left(1 - \frac{S_k}{S_{max}} \right) \quad (4)$$

where I and E are the maximal immigration and emigration rates, respectively. The pseudo code of the migration operator is depicted in Algorithm 1.

Consider the special case $E = I$; (3) and (4) can be combined as

$$\mu_k + \lambda_k = E \quad (5)$$

2.2.2. Mutation. In the BBO algorithm, some events, such as the wind carrying seeds or flotsam, provide more favorable features that allow an island to generate superior solutions with statistically significant enhancements. Through species count probabilities P_s , the mutation rate m_i can be determined as

$$m_i = m_{max} \left(1 - \frac{P_s}{P_{max}} \right) \quad (6)$$

where m_{max} is a user-defined maximal mutation rate that m can reach, and $P_{max} = \max(P_s)$. The mutation scheme tends to increase diversity among the population. Highly probable solutions tend to be more dominant in the population. Thus, the high HSI solutions likely mutate, which increases the probability that they will improve even more than they already have. The pseudo code of the mutation operator is depicted in Algorithm 2.

The BBO algorithm can be described using the following steps:

- (1) Initialize the BBO parameters, which include a problem-dependent method for mapping problem solutions to SIVs and habitats, the modification probability P_{mod} , the maximal species count S_{max} , the maximal migration rates E and I , the maximal mutation rate m_{max} , and elite number p .

```

1: For each habitat
2:   For each SIV
3:   Select habitat  $H_i$  based on the mutation probability  $m_i$ .
4:   If  $U(0,1) < m_i$  then
5:     Replace  $H_i(SIV)$  by a randomly generated SIV.
6:   End
7: End
8: End

```

ALGORITHM 2: Pseudo code for mutation of BBO.

- (2) Initialize the habitats, which depend on the population size and problem to determine what habitat is a potential solution.
- (3) Calculate the HSI of each habitat, emigration rate μ , immigration rate λ , and species S .
- (4) Identify the elite habitats depending on the HSI value.
- (5) Migration operation: Modify each nonelite habitat by immigration and emigration rates. The probability that a habitat H_i is modified is proportional to its immigration rate λ_i , and the probability that the source of the modification comes from a habitat H_j is proportional to the emigration rate μ_j . After modification of each nonelite habitat using the migration operation, each HSI is recomputed.
- (6) Update the species count probability within the habitat according to (2). The mutation operation is performed on each nonelite habitat, and compute each HSI value.
- (7) Go to step (3) for the next iteration. The BBO operation is terminated if the criteria are satisfied.

The BBO parameters, habitats, and HSI value evaluations for identifying the relationship between the clinicopathologic variables of breast cancer are explained in Section 2.2.4.

2.2.3. Initializing the BBO Parameters and Habitats. In this study, the BBO algorithm parameters were set as follows: habitat size = 50, number of generation = 100, maximum immigration and emigration rates for each island = 1, elite number = 2, and mutation probability = 0.04 [30].

The solution H comprises h_i ($i = 1$ to population number) as in (7), each habitat h consists of SIV_n ($n = 1$ to number of problem dimensions) as in (8), and each SIV_n consists of a randomly generated 0 or 1 as in (9).

$$H = \{h_1, h_2, \dots, h_i\} \quad (7)$$

$$h = \{SIV_1, SIV_2, \dots, SIV_n\} \quad (8)$$

$$SIV_j \in \{0, 1\}, \quad j = 1, 2, 3, \dots, n \quad (9)$$

2.2.4. Evaluation of HSI Values. We use the C-statistic (or “concordance” statistic) as the HSI values in the BBO algorithm; the C-statistic value is calculated according to linear prediction estimated through Cox proportional hazard (PH)

regression. The C-statistic's goodness of fit can be measured for binary outcomes in a regression model. The value of the C-statistic indicates the probability that a patient who had experienced an event had a higher mortality risk than one who had not experienced the event. C-statistics were used to test the predictive accuracy of survival models. The value of the C-statistic equals the area under the receiver operating characteristic curve. A C-statistic value of 0.5 indicates that a model predicts the outcome by chance, 0.7–0.8 indicates acceptable discrimination, 0.8–0.9 indicates excellent discrimination, 0.9–0.99 indicates outstanding discrimination, and 1.0 is perfect prediction [34].

Harrell [35] defined the C-statistic as the ratio of all sample pairs for which the predictions and real outcomes are concordant. Assume a sample data of M items is given. Let $X_1, X_2, X_3, \dots, X_M$ indicate the survival times, and $Y_1, Y_2, Y_3, \dots, Y_M$ indicate the predicted probabilities of survival. Given that $X_i \neq X_j$, each concordant pair is assigned 1 point, and each discordant pair is assigned 0 point [36].

$$c_{ij} = \begin{cases} 1, & \text{if } X_i < X_j \text{ and } Y_i < Y_j \text{ or } X_i > X_j \text{ and } Y_i > Y_j \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

Subsequently, the C-statistic is calculated using the following equation:

$$C = \frac{1}{u_{ij(i,j) \in U}} \sum c_{ij} \quad (11)$$

where u_{ij} is the number of all usable pairs, and U is a set of all usable pairs of participants (i, j) .

2.3. Statistical Analysis. The differences in the distribution of the clinicopathologic variables between surviving and deceased participants were estimated using chi-squared tests. A univariate Cox PH regression model was used to evaluate the hazard function of each clinicopathologic variable in the overall survival of patients with breast cancer. A fully adjusted multivariate Cox PH model was used to estimate the association of all clinicopathologic variables in overall survival among patients with breast cancer. We applied the three stepwise selection Cox PH models by adjusting the selective criteria where $p < 0.05, 0.1,$ and 0.2 for clinicopathologic variable inclusion. By contrast, the BBO–Cox PH model included only the clinicopathologic variables selected using the BBO algorithm. The performance of each Cox PH multivariate model was determined using the C-statistics value. The hazard ratio (HR), 95% confidence interval (95% CI), and p -value were computed. The overall survival function of the statistically significant variables in the selected BBO model was visualized using a Kaplan–Meier curve to establish the effects of each individual clinicopathologic variable on overall survival. The differences in overall survival function between different strata were tested using the log-rank test; $p = 0.05$ indicated statistical significance for all results of statistical analysis. The statistics were analyzed using SAS 9.3 software (SAS Institute Inc., Cary, NC, USA).

3. Results

3.1. Clinicopathologic Characteristics and Survival of Patients with Breast Cancer. A comparison of the clinicopathologic variables between surviving and deceased patients is presented in Table 1; it shows that 1830 (96.52%) and 66 (3.48%) of the 1896 patients with breast cancer survived and died, respectively. Compared with the survival group, a greater proportion of the death group was HER2 positive ($p = 0.040$), ER positive ($p < 0.001$), PR positive ($p = 0.001$), with tumor size ≥ 2.0 cm ($p < 0.001$), lymph node positive ($p < 0.001$), LVI positive ($p < 0.001$), with dermal invasion ($p = 0.009$), with perineural invasion ($p < 0.001$), with total mastectomy ($p < 0.001$), without chemotherapy treatment ($p = 0.004$), and without hormone therapy ($p < 0.001$).

3.2. Univariate Clinicopathologic Variables. The 15 clinicopathologic variables included age, grade, tumor size, ER, PR, HER2, lymph node, LVI, dermal invasion, perineural invasion, surgery method, radiotherapy, chemotherapy, hormone therapy, and target therapy. Each clinicopathologic variable is dichotomized into low- and high-risk characteristics in Table 2, which lists the clinical characteristics for low- and high-risk breast cancer progression; these characteristics were identical to those used in previous studies. The univariate Cox regression analysis of the clinicopathologic variables in overall survival of patients with breast cancer is presented in Table 2. Grade (HR = 2.169, $p = 0.002$), ER (HR = 0.4, $p < 0.001$), PR (HR 0.494, $p = 0.005$), tumor size (HR = 4.362, $p < 0.001$), lymph node (HR = 2.689, $p < 0.001$), LVI (HR = 2.960, $p < 0.001$), dermal invasion (HR = 3.060, $p = 0.003$), perineural invasion (HR = 2.558, $p < 0.001$), surgical method (HR = 2.981, $p < 0.001$), and hormone therapy (HR = 2.666, $p < 0.001$) were significantly associated with breast cancer progression, including mortality and disease progression.

3.3. Multivariate Clinicopathologic Variables. We compared the fully adjusted Cox PH model, three stepwise selection Cox PH models with different p -value inclusion levels (0.05, 0.1, and 0.2), and BBO–Cox PH model. The performance of each model is summarized in Table 3. The fully adjusted Cox PH model included all 15 clinicopathologic variables. The fully adjusted Cox PH model indicated that older age (adjusted HR = 0.515, $p = 0.015$) was correlated with a high risk of mortality for patients with breast cancer, whereas tumor sizes greater than 2 cm (adjusted HR = 2.430, $p = 0.005$) and LVI (adjusted HR = 1.776, $p = 0.046$) were significantly correlated with a high risk of breast cancer progression.

In the stepwise selection Cox PH model with inclusion at $p = 0.05$, tumor sizes greater than 2 cm (adjusted HR = 3.017, $p < 0.005$), LVI (adjusted HR = 1.878, $p = 0.016$), dermal invasion (adjusted HR = 2.152, $p = 0.048$), and without hormone therapy (adjusted HR = 2.231, $p = 0.002$) were significantly correlated with a high risk of breast cancer progression. In the stepwise selection Cox PH model with inclusion at $p = 0.1$, older age (adjusted HR = 0.589, $p = 0.036$) was correlated with a high risk of mortality for patients with breast cancer, whereas grade III (adjusted HR = 1.830, $p =$

TABLE 1: Comparison of clinicopathologic variables between the surviving and deceased patients.

Variable	Total		Survival		Expired		<i>p</i> -value*
	N	%	N	%	N	%	
Age (years)							0.123
≤ 50	719	37.92	688	37.60	31	46.97	
> 50	1177	62.08	1142	62.40	35	53.03	
Grade							0.122
I, II	1286	67.83	1247	68.14	39	59.09	
III	610	32.17	583	31.86	27	40.91	
Her 2							0.040
Negative	1230	64.87	1195	65.30	35	65.30	
Positive	666	35.13	635	34.70	31	46.97	
Estrogen receptor							<0.001
Negative	510	26.90	476	26.01	34	51.52	
Positive	1386	73.10	1354	73.99	32	48.48	
Progesterone receptor							0.001
Negative	701	36.97	664	36.28	37	56.06	
Positive	1195	63.03	1166	63.72	29	43.94	
Tumor size (AJCC stage)							<0.001
≤ 2 cm	1161	61.23	1142	62.40	19	28.79	
> 2 cm	735	38.77	688	37.60	47	71.21	
Lymph node (AJCC stage)							<0.001
Negative	1265	33.72	1240	67.67	25	37.88	
Positive	631	33.28	590	32.24	41	62.12	
Lymphovascular invasion							<0.001
Negative	1326	69.94	1293	70.66	33	50.00	
Positive	570	30.06	537	29.34	33	50.00	
Dermal invasion							0.009
Negative	1798	94.83	1740	95.08	58	87.88	
Positive	98	5.17	90	4.92	8	12.12	
Perineural invasion							<0.001
Negative	1659	87.5	1610	87.98	49	74.24	
Positive	237	12.50	220	12.02	17	25.76	
Surgery method							<0.001
Partial mastectomy	1133	59.76	1113	60.82	20	30.30	
Total mastectomy	763	40.24	717	39.18	46	69.70	
Radiotherapy (RT)							0.050
With RT	1194	62.97	1160	63.39	34	51.52	
Without RT	702	37.03	670	36.61	32	48.48	
Chemotherapy							0.004
With CT	1140	60.13	1089	59.51	51	77.27	
Without CT	756	39.87	741	40.49	15	22.73	
Hormone therapy							<0.001
With HT	1265	69.13	28	42.42	1265	69.13	
Without HT	565	30.87	38	57.58	565	30.87	
Target therapy							0.232
With TT	302	15.93	288	15.74	14	21.21	
Without TT	1594	84.07	1542	84.26	52	78.79	

**p*-value is estimated using the chi-squared test.
HER2: human epidermal growth factor receptor 2.

TABLE 2: Univariate Cox regression analysis of clinicopathologic variables in breast cancer overall survival.

Variable	Comparison	HR	Univariate 95% CI	p-value*
Age	>50 years	0.789	0.486-1.281	0.338
Grade	III	2.169	1.332-3.558	0.002
HER 2	Positive	1.036	0.638-1.683	0.887
Estrogen receptor	Positive	0.400	0.246-0.650	<0.001
Progesterone receptor	Positive	0.494	0.304-0.804	0.005
Tumor size (AJCC stage)	≥ 2.4 cm	4.362	2.556-7.448	<0.001
Lymph node (AJCC stage)	Positive	2.689	1.631-4.434	<0.001
Lymphovascular invasion	Positive	2.960	1.825-4.801	<0.001
Dermal invasion	Positive	3.060	1.453-6.445	0.003
Perineural invasion	Yes	2.558	1.473-4.444	<0.001
Surgery method	Total mastectomy	2.981	1.761-5.047	<0.001
Radiotherapy (RT)	Without RT	1.446	0.891-2.348	0.135
Chemotherapy (CT)	Without CT	0.782	0.438-1.397	0.406
Hormone therapy (HT)	Without HT	2.666	1.636-4.345	<0.001
Target therapy (TT)	Without TT	0.779	0.432-1.704	0.409

*p-value is estimated using the Cox PH regression.

HR: hazard ratio; HER2: human epidermal growth factor receptor 2.

TABLE 3: Multivariate Cox regression analysis of clinicopathologic variables in breast cancer overall survival.

Variable	Comparison		Fully adjusted		Model 1 ¹		Stepwise selected Model 2 ²		Model 3 ³		BBO selected	
	>50 years	≤50 years	HR	p-value*	HR	p-value*	HR	p-value*	HR	p-value*	HR	p-value*
C-statistic												
p-value*												
Age												
Grade	III	I, II	1.524	0.132	1.830	0.026	1.642	0.073				
Her 2	Positive	Negative	0.103	0.103								
Estrogen receptor	Positive	Negative	0.506	0.195			0.451	0.003				
Progesterone receptor	Positive	Negative	0.801	0.595								
Tumor size (AJCC stage)	≥ 2.4 cm	< 2.4 cm	2.430	0.005	3.017	<0.001	2.435	0.004	2.436	0.004	2.372	0.006
Lymph node (AJCC stage)	Positive	Negative	1.511	0.103							1.301	0.038
Lymphovascular invasion	Positive	Negative	1.776	0.046	1.878	0.016	1.642	0.068	1.753	0.038	1.606	0.096
Dermal invasion	Positive	Negative	1.549	0.298	2.152	0.048			1.722	0.188	1.548	0.028
Perineural invasion	Yes	No	1.745	0.081			1.994	0.028	1.827	0.060	1.546	0.168
Surgery method	Total mastectomy	Partial mastectomy	1.735	0.087			1.869	0.031	1.848	0.066	1.633	0.092
Radiotherapy	Without RT	With RT	1.299	0.378							1.234	0.491
Chemotherapy	Without CT	With CT	1.338	0.379							2.178	0.003
Hormone therapy	Without HT	With HT	1.022	0.962	2.231	0.002	1.918	0.014				
Target therapy	Without TT	With TT	0.859	0.697								

¹ Stepwise selection met the $p = 0.05$ level for entry into the model; ² Stepwise selection met the $p = 0.1$ level for entry into the model; ³ Stepwise selection met the $p = 0.2$ level for entry into the model.

*p-value is estimated using Cox PH regression.

HR: hazard ratio; HER2: human epidermal growth factor receptor 2.

0.026), tumor sizes greater than 2 cm (adjusted HR = 2.435, $p = 0.004$), perineural invasion (adjusted HR = 1.994, $p = 0.028$), total mastectomy (adjusted HR 1.869, $p = 0.031$), and without hormone therapy (adjusted HR = 1.918, $p = 0.014$) were significantly correlated with a high risk of breast cancer progression. In the stepwise selection Cox PH model with inclusion at $p = 0.2$, age (adjusted HR = 0.523, $p = 0.003$), PR (adjusted HR = 0.451, $p = 0.003$), tumor size (adjusted HR = 2.436, $p = 0.004$), and LVI (adjusted HR = 1.753, $p = 0.038$) were significantly correlated with breast cancer progression.

The BBO-Cox PH model included tumor size, lymph node, dermal invasion, surgery method, chemotherapy, and hormone therapy as breast cancer progression predictors. Tumor sizes greater than 2 cm (HR = 2.372, $p = 0.006$), lymph node metastasis (HR = 1.301, $p = 0.038$), dermal invasion (HR = 1.548, $p = 0.028$), and without hormone therapy (HR = 1.178, $p = 0.003$) were significantly correlated with breast cancer progression.

The C-statistic values were 76%, 78%, 78%, 77%, and 80% in the fully adjusted Cox PH model, three stepwise selection Cox PH (with inclusion at $p = 0.05$, 0.1, and 0.2) models, and BBO-Cox PH model, respectively. The BBO-Cox PH model obtained the highest C-statistic value of all the models, which indicated higher concordance for predicting the overall survival of patients with breast cancer.

We present the survival curves of various clinicopathologic characteristic categories using Kaplan-Meier method. The difference between each clinicopathologic variable in the survival curves was estimated using the log-rank test. The BBO-Cox PH model included tumor size, lymph node, dermal invasion, surgery method, chemotherapy, and hormone therapy as breast cancer progression predictors, and the results revealed that all BBO-selected clinicopathologic variables were statistically significant in different categories within each individual clinicopathologic variable. The Kaplan-Meier curves are presented in Figure 2.

4. Discussion

Overfitting is a limitation that must be faced for conventional statistical approaches, such as regression models. The principal reason is that regression-based selection approaches provide analysis results that depend mainly on the distribution of a population in associated variables. BBO is similar to evolutionary algorithms, such as the genetic algorithm (GA) [37] and particle swarm optimization (PSO) [38]. GA, PSO, and BBO are all inspired by nature and perform information sharing between solutions (i.e., genes for GA, particles for PSO, and habitats for BBO) [39]. A GA solution will “die” at the end of each generation, and PSO and BBO solutions will continue to exist as the optimization process progresses. PSO solutions tend to be clustered together in similar groups, whereas GA and BBO solutions do not necessarily have any clustering tendencies. BBO has unique features, and a BBO solution shares its content (SIV) directly with other solutions. This feature leads to BBO having better performance than GA

and PSO in optimization problems [39]. BBO has been successfully applied to several engineering problems, including global benchmark functions and economic load scheduling [33]. In this study, BBO was successfully applied to select clinicopathologic variables that resulted in superior performance compared with other Cox regression models (i.e., the fully adjusted Cox regression model and three stepwise-selected Cox regression models). BBO is a mechanism of the Cox regression model that improves the ability to explore C-statistics and variables. By using BBO, the migration operator shared more accurate information than with other solutions, and including randomization with the mutation operator at the mutation stage partly resolved the overfitting problem.

The BBO-selected model included tumor size, lymph node metastasis, lymphovascular invasion, perineural invasion, dermal invasion, surgery method, hormone therapy, and chemotherapy as predictors for the overall survival of patients with breast cancer. This optimization model obtained superior prediction results and concordance compared with the fully adjusted Cox PH model in the assessment of C-statistics. The fact that the BBO algorithm considers the internal relationship between clinicopathologic variables may explain why it increases the discriminant ability of the final BBO model.

In addition, we reevaluated the individual HR for each selected variable in the BBO model through multivariate Cox PH regression analysis. The results revealed that not all of the selected variables provided a significant individual HR in the Cox PH model. Only tumor size, lymph node metastasis, dermal invasion, and hormone therapy had a significant correlation with the overall survival of patients with breast cancer. These factors had been previously reported as associated with the overall survival of patients with breast cancer.

Tumor size and lymph node are frequently used to predict the overall survival and prognosis of patients with breast cancer [40]. The NPI index is a benchmark for breast cancer prognoses that includes these two main factors [13, 41]. The extended use of the NPI index for various breast cancer prognostic purposes remained contributed and widespread in recent studies such as including the hormone status (ER, PR, and Her2) [42–44]. Dermal invasion, also known as dermal lymphatic invasion, is generally associated with inflammatory breast cancer [45]. Breast cancer characterized by dermal invasion is generally correlated with inferior overall survival. Hormone therapy is an effective therapy for hormone-positive patients with breast cancer [46]. Therefore, the decision to commence hormone treatment mainly relies on the ER and PR status of patients with breast cancer. Although ER and PR were not included in the BBO-selected model, hormone therapy (an effective treatment options for patients with ER positive or PR positive breast cancer) could partially explain the beneficial effect of ER and PR positive case with hormone therapy in breast cancer overall survival. However, the retrospective nature of our study might have limited the analysis of various covariates that have been previously reported as correlated with overall breast cancer survival rates.

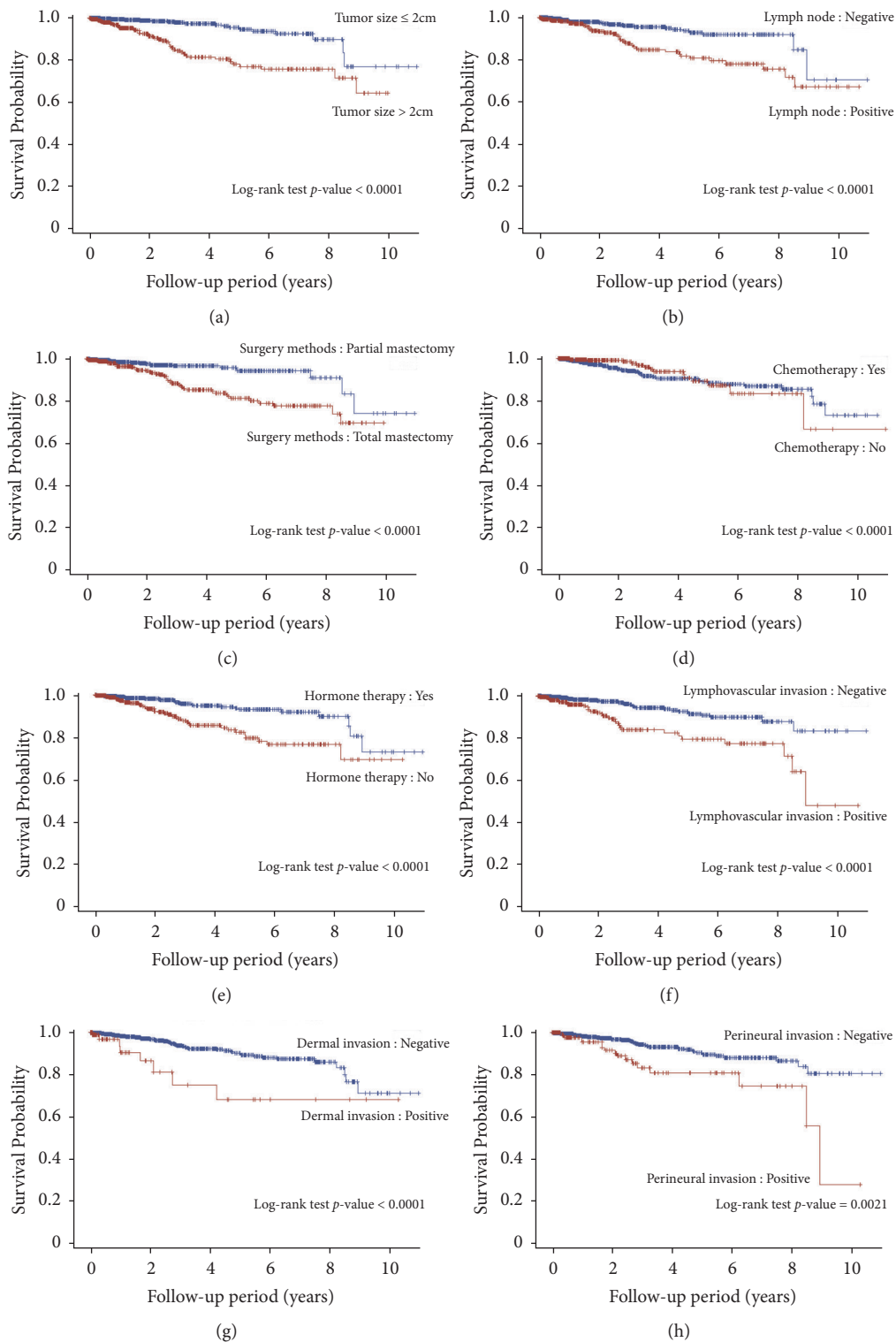


FIGURE 2: Kaplan–Meier curve of BBO-selected clinicopathologic variables including (a) tumor size, (b) lymph node status, (c) lymphovascular invasion status, (d) dermal invasion status, (e) perineural invasion status, (f) surgical methods, (g) chemotherapy status, and (h) hormone therapy status.

5. Conclusion

This study used the BBO algorithm to search for combinations of clinicopathologic variables that can facilitate prediction of the overall survival of patients with breast cancer. Compared with the fully adjusted multivariate Cox regression model and stepwise selection Cox regression model, the BBO-selected model had a higher C-statistic value for predicting overall survival. This study determined that the BBO algorithm could select only 8 variables from 15 clinicopathologic variables, which is the minimum number of discriminators required for optimal discriminant effectiveness when predicting the overall survival of patients with breast cancer accurately and with concordance. This model may have vital implications for the selection of clinicopathologic variables for breast cancer.

Data Availability

The data used to support the findings of this study have not been made available because of the institution regulation.

Conflicts of Interest

The authors declare that they have no competing interests or other financial motivations.

Acknowledgments

This work was partly supported by the Ministry of Science and Technology (105-2221-E-151-053-MY2, 107-2811-E-992-500-, 105-2314-B-037-038-MY3, and 105-2314-B-037-037-MY3), Taiwan.

References

- [1] M. o. H. a. W. Health Promotion Administration, "Cancer registry annual report, 2015," 2017.
- [2] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA: A Cancer Journal for Clinicians*, vol. 61, no. 2, pp. 69–90, 2011.
- [3] N. J. Bundred, "Prognostic and predictive factors in breast cancer," *Cancer Treatment Reviews*, vol. 27, no. 3, pp. 137–142, 2001.
- [4] O. Y. Fu, Y. D. Lin, L. Y. Chuang, H. W. Chang, C. H. Yang, and M. F. Hou, "The combinational polymorphisms of ORAI1 gene are associated with preventive models of breast cancer in the Taiwanese," *Biomed Research International*, vol. 2015, Article ID 281263, 7 pages, 2015.
- [5] L.-Y. Chuang, Y.-D. Lin, H.-W. Chang, and C.-H. Yang, "An improved PSO algorithm for generating protective SNP barcodes in breast cancer," *PLoS ONE*, vol. 7, no. 5, Article ID e37018, 2012.
- [6] W.-C. Chang, Y.-Y. Fang, H.-W. Chang et al., "Identifying association model for single-nucleotide polymorphisms of ORAI1 gene for breast cancer," *Cancer Cell International*, vol. 14, no. 1, article 29, 2014.
- [7] C.-H. Yang, Y.-D. Lin, L.-Y. Chuang, and H.-W. Chang, "Double-bottom chaotic map particle swarm optimization based on chi-square test to determine gene-gene interactions," *BioMed Research International*, vol. 2014, Article ID 172049, 10 pages, 2014.
- [8] C.-H. Yang, Y.-D. Lin, L.-Y. Chuang, and H.-W. Chang, "Evaluation of breast cancer susceptibility using improved genetic algorithms to generate genotype SNP barcodes," *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 10, no. 2, pp. 361–371, 2013.
- [9] O.-Y. Fu, H.-W. Chang, Y.-D. Lin, L.-Y. Chuang, M.-F. Hou, and C.-H. Yang, "Breast cancer-associated high-order SNP-SNP interaction of CXCL12/CXCR4-related genes by an improved multifactor dimensionality reduction (MDR-ER)," *Oncology Reports*, vol. 36, no. 3, pp. 1739–1747, 2016.
- [10] T. Decker, D. Hungermann, and W. Böcker, "Prognostische und prädiktive Faktoren invasiver Mammakarzinome," *Der Pathologe*, vol. 30, no. 1, pp. 49–55, 2009.
- [11] E. M. Ibrahim, M. E. Al-Foheidi, M. M. Al-Mansour, and G. A. Kazkaz, "The prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancer: a meta-analysis," *Breast Cancer Research and Treatment*, vol. 148, no. 3, pp. 467–476, 2014.
- [12] K. N. Stevens, C. M. Vachon, and F. J. Couch, "Genetic susceptibility to triple-negative breast cancer," *Cancer Research*, vol. 73, no. 7, pp. 2025–2030, 2013.
- [13] T. J. Anderson, "Breast cancer prognostication in the 21st century and the Nottingham prognostic index," *Journal of Clinical Pathology*, vol. 55, no. 2, pp. 86–87, 2002.
- [14] Y. Fong, J. Evans, D. Brook, J. Kenkre, P. Jarvis, and K. Gower-Thomas, "The nottingham prognostic index: five- and ten-year data for all-cause survival within a screened population," *Annals of the Royal College of Surgeons of England*, vol. 97, no. 2, pp. 137–139, 2015.
- [15] R. Jimenez-Lee, B. Ham, J. Vetto, and R. Pommier, "Breast cancer severity score is an innovative system for prognosis," *The American Journal of Surgery*, vol. 186, no. 4, pp. 404–408, 2003.
- [16] J. Brandt, J. Garne, I. Tengrup, and J. Manjer, "Age at diagnosis in relation to survival following breast cancer: a cohort study," *World Journal of Surgical Oncology*, vol. 13, no. 1, p. 33, 2015.
- [17] A. H. Partridge, M. E. Hughes, R. A. Ottesen et al., "The effect of age on delay in diagnosis and stage of breast cancer," *The Oncologist*, vol. 17, no. 6, pp. 775–782, 2012.
- [18] A. Purushotham, E. Shamil, M. Cariati et al., "Age at diagnosis and distant metastasis in breast cancer - a surprising inverse relationship," *European Journal of Cancer*, vol. 50, no. 10, pp. 1697–1705, 2014.
- [19] X. Yu, S. Zhou, J. Wang et al., "Hormone replacement therapy and breast cancer survival: a systematic review and meta-analysis of observational studies," *Breast Cancer*, vol. 24, no. 5, pp. 643–657, 2017.
- [20] E. C. Inwald, O. Ortmann, F. Zeman et al., "Guideline concordant therapy prolongs survival in HER2-positive breast cancer patients: results from a large population-based cohort of a cancer registry," *BioMed Research International*, vol. 2014, Article ID 137304, 10 pages, 2014.
- [21] S. B. L. Makanjuola, A. O. Popoola, and M. A. Oludara, "Radiation therapy: a major factor in the five-year survival analysis of women with breast cancer in Lagos, Nigeria," *Radiotherapy & Oncology*, vol. 111, no. 2, pp. 321–326, 2014.
- [22] B. Chang, J. Sokhn, E. James, and M. Abu-Khalaf, "Prolonged progression-free survival in a patient with triple-negative breast cancer metastatic to the liver after chemotherapy and local

- radiation therapy," *Clinical Breast Cancer*, vol. 14, no. 2, pp. e61–e64, 2014.
- [23] A. A. Sofi, I. Mohamed, M. Koumaya, and Z. Kamaluddin, "Local therapy in metastatic breast cancer is associated with improved survival," *American Journal of Therapeutics*, vol. 20, no. 5, pp. 487–492, 2013.
- [24] M. D. Ritchie, L. W. Hahn, N. Roodi et al., "Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer," *American Journal of Human Genetics*, vol. 69, no. 1, pp. 138–147, 2001.
- [25] M. Montazeri, M. Montazeri, M. Montazeri, and A. Beigzadeh, "Machine learning models in breast cancer survival prediction," *Technology and Health Care*, vol. 24, no. 1, pp. 31–42, 2016.
- [26] L. Vanneschi, A. Farinaccio, G. Mauri, M. Antoniotti, P. Provero, and M. Giacobini, "A comparison of machine learning techniques for survival prediction in breast cancer," *BioData Mining*, vol. 4, no. 12, 2011.
- [27] A. Modiri and K. Kiasaleh, "Permittivity estimation for breast cancer detection using particle swarm optimization algorithm," in *Proceedings of the 2011 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1359–1362, Boston, Mass, USA, August 2011.
- [28] D. Simon, "Biogeography-based optimization," *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 6, pp. 702–713, 2008.
- [29] A. A. Christy and P. A. D. V. Raj, "Adaptive biogeography based predator-prey optimization technique for optimal power flow," *International Journal of Electrical Power & Energy Systems*, vol. 62, pp. 344–352, 2014.
- [30] H. Ma and D. Simon, "Blended biogeography-based optimization for constrained optimization," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 3, pp. 517–525, 2011.
- [31] Y.-J. Zheng, H.-F. Ling, X.-B. Wu, and J.-Y. Xue, "Localized biogeography-based optimization," *Soft Computing*, vol. 18, no. 11, pp. 2323–2334, 2014.
- [32] Y.-J. Zheng, H.-F. Ling, and J.-Y. Xue, "Ecogeography-based optimization: enhancing biogeography-based optimization with ecogeographic barriers and differentiations," *Computers & Operations Research*, vol. 50, pp. 115–127, 2014.
- [33] H. P. Ma, D. Simon, P. Siarry, and M. R. Fei, "Biogeography-based optimization: a 10-year review," *IEEE Transactions on Emerging Topic in Computational Intelligence*, vol. 1, no. 5, pp. 391–407, 2017.
- [34] D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, John Wiley & Sons, 2013.
- [35] F. E. Harrell Jr., K. L. Lee, and D. B. Mark, "Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors," *Statistics in Medicine*, vol. 15, no. 4, pp. 361–387, 1996.
- [36] M. J. Pencina and R. B. Agostino, "Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation," *Statistics in Medicine*, vol. 23, no. 13, pp. 2109–2123, 2004.
- [37] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, MIT Press, 1992.
- [38] R. C. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proceedings of the 6th International Symposium on Micromachine and Human Science*, pp. 39–43, Nagoya, Japan, October 1995.
- [39] N. Hordri, S. Yuhaniz, and D. Nasien, "A comparison study of biogeography based optimization for optimization problems," *International Journal of Advances in Soft Computing and its Applications*, vol. 5, pp. 1–16, 2013.
- [40] L. Cortesi, L. Marcheselli, V. Guarneri et al., "Tumor size, node status, grading, HER2 and estrogen receptor status still retain a strong value in patients with operable breast cancer diagnosed in recent years," *International Journal of Cancer*, vol. 132, no. 2, pp. E58–E65, 2013.
- [41] A. H. Lee and I. O. Ellis, "The nottingham prognostic index for invasive carcinoma of the breast," *Pathology & Oncology Research*, vol. 14, no. 2, pp. 113–115, 2008.
- [42] K.-J. Winzer, A. Buchholz, M. Schumacher, and W. Sauerbrei, "Improving the prognostic ability through better use of standard clinical data - the nottingham prognostic index as an example," *PLoS ONE*, vol. 11, no. 3, Article ID e0149977, 2016.
- [43] H. Zhen, L. Yang, L. Li et al., "Correlation analysis between molecular subtypes and nottingham prognostic index in breast cancer," *Oncotarget*, vol. 8, no. 43, pp. 74096–74105, 2017.
- [44] J. Kwon, K.-Y. Eom, T. R. Koo et al., "A prognostic model for patients with triple-negative breast cancer: importance of the modified nottingham prognostic index and age," *Journal of Breast Cancer*, vol. 20, no. 1, pp. 65–73, 2017.
- [45] M. C. Abramowitz, T. Li, M. Morrow et al., "Dermal lymphatic invasion and inflammatory breast cancer are independent predictors of outcome after postmastectomy radiation," *American Journal of Clinical Oncology*, vol. 32, no. 1, pp. 30–33, 2009.
- [46] T. Reinert and C. H. Barrios, "Overall survival and progression-free survival with endocrine therapy for hormone receptor-positive, HER2-negative advanced breast cancer: review," *Therapeutic Advances in Medical Oncology*, vol. 9, no. 11, pp. 693–709, 2017.