STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **13**, May 2025, pp 2105–2118.
Published online in International Academic Press (www.IAPress.org)

Variable Selection in Weibull Accelerated Survival Model Based on Chaotic Sand Cat Swarm Algorithm

Ahmed Naziyah alkhateeb [1,*], Qutaiba N. Nayef Al-Qazaz [2]

[1] *Department of Operation Research and Intelligent Techniques, University of Mosul, Iraq*
[2] *Department of Statistics, University of Baghdad, Iraq*

Abstract In medical research, proportional hazard models are much more common, but accelerated failure time (AFT) models are still widely used. Variables in the AFT model influence the event time by altering the logarithm of the dependent variable's survival time. The parametric forms typically utilized by AFT models are restricted and cannot be represented otherwise. The selection of variables and parameter estimation for the Weibull distribution is a common practice. This predictive approach is often applied in reliability studies in engineering and medical forecasts, particularly for estimating survival time. Additionally, we present an empirical example using our prediction method on a publicly accessible dataset. Sand cat swarm optimization (SCSO) is a new metaheuristic algorithm that mimics the survival behavior of sand cats. The results reveal that SCSO outperforms other methods in terms of convergence speed and finds all or most local/global optima. The SCSO algorithm is introduced to identify critical variables in the Weibull AFT regression model. Thus, variations of the SCSO algorithm can be applied to address the Weibull AFT problem.

Keywords Weibull distribution, High dimensional, Accelerated Failure Time model, Feature Selection, Survival Analysis, SCSO

AMS 2010 subject classifications 60E05, 62E15

DOI: 10.19139/soic-2310-5070-2077

1. Introduction

survival data is information sets that quantify the time until an event occurs. Some interesting variables include the employment duration and heart transplant recipients' survival rate. Any analysis involving such data must consider certain factors [1, 2].

An event time variable that indicates how long it will take for a particular event to occur and makes up survival data, as does a potential collection of independent variables that are hypothesized to be related to the failure time variable. These independent variables, also known as concurrent variables, covariates, or prognostic factors, might be continuous, like age or temperature, or discrete, like sex or race. As with most medical data, the system that causes the event of interest might be biological or physical, as in the case of engineering data [3, 4]. The two primary goals of survival analysis are modeling the failure time variable's underlying distribution and evaluating its reliance on the independent variables.

In both cases, the only rate of failure that is clearly apparent to be related to a given censored observation is the lower boundary. It postulated that such observation is right censored. An extra variable is added to the study to identify which failure times are censored and which are observed event times. More broadly, the failure time may only be known to be inside a specific interval (interval-censored) or to be smaller than a set value

*Correspondence to: Ahmed Naziyah alkhateeb (Email: ahmed.alkhateeb@uomosul.edu.iq). Department of Operation Research and Intelligent Techniques, College of Computer Sciences and Mathematics, University of Mosul, Nineveh, Iraq.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

(left-censored). In survival analysis, several potential censorship strategies surface. In addition to [5] discussing several filtering systems, [6] examines some comparable censoring circumstances. Longer-lived persons are often more likely to be right-censored, among other reasons; hence, data including censored observations cannot be analyzed, disregarding the censored observations. The censoring must be considered in the analysis process, and the censored and uncensored observations must be used appropriately. [2] applied AFT to patients with acute liver failure in India, AFT survival models can provide the appropriate model for data if the researcher can determine an appropriate survival time distribution.

AFT is a fully parametric model, allowing it to make conclusions that would be challenging in a non- or semi-parametric framework, such as estimating tail probabilities. AFT models are linear mixed models that analyze survival time data by applying a log transformation, accounting for censoring and capturing the correlation in survival data. The compromise is that one must assume a particular survival time distribution, which may need to be revised. According to the AFT model, a covariate's impact will either speed up or slow down a disease's course by a certain amount. In AFT models, a covariate affects the full distribution of the response variable and the time scale. On the other hand, in the Proportional Hazard (PH) models, the effect of covariate is in multiplicative sense in terms of hazard rates [7]. Parametric survival models exist that do not necessitate the restricted assumption of PH. Accelerated failure time (AFT) models provide benefits in interpretability, management of censored data, and versatility in mediation analysis with survival results compared to proportional hazards (PH) models. Researchers should take into account these distinctions when selecting the suitable model for their mediation analysis [4, 8]. Conditional models of Cox type and models of AFT are based on some basic assumptions that are the basis of the connection between the survival and coincidence of the variables. Another common and rather flexible model is the Cox model, which can be used assuming that the underlying assumption of proportional hazards is met. The PH assumption requires all independent variables to be independent of time in the final model for the Cox model to be used appropriately; in other words, the risk ratio of the event remains constant, although it happens over a given period. The precision of this assumption will facilitate further investigation of the output classification process, which will be more facilitated compared to the interpretation of parametric models [9]. AFT models are beneficial for analyzing the distribution of survival times and are particularly helpful when the proportional hazards assumption is not met [10]. Furthermore, parametric survival models offer several benefits, including the ability to use complete likelihood to estimate parameters and provide estimates based on survival rather than outcome hazards [2]. [11] recommend using AFT models as the recommended analytical tool for analyzing latency to the platform in Multi-Walled Carbon Nanotubes on Morris water maze research by comparing AFT with the logistic regression model and analysis of variance (ANOVA).

The Weibull distribution [12], whose application was started within the industrial field to carry out reliability testing, is one of the most obvious probabilistic distributions capable of studying survival data. This distribution is just as necessary to the parametric analysis of survival data as linear modeling's normal distribution. The AFT assumption is suitable for comparing survival times, while the PH assumption is suitable for comparing hazards [13].

2. Regression Models for Survival Data

We need alternative regression models because using a linear regression model is not a feasible solution given the nature of the data. Logistic regression models are used for binary outcomes that are nominal or ordinal. Because baseline risks or survivor functions are not stated, proportional hazards and accelerated failure models are semiparametric models. However, if the underlying probability distributions are provided, the accelerated models can also be parametric [14].

Regression models for survival data are used to statistically analyze data in which the outcome variable is the duration of time until a specific event occurs. The methodology of survival analysis can be extended to several alternative scenarios, but common events include failure, relapse, and death. For survival data, the Cox Proportional Hazards model and parametric survival models are the two main categories of regression models. The proportional

hazard rate (PH) model is defined as follows:

$$h(tX_i) = h_0(t).\exp(\beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p) \tag{1}$$

Where $h_0(t)$ are defined as baseline hazard, $\beta_1, \beta_2, ...\beta_p$ are coefficient regression of the independent variable $X_1, X_2, ..., X_p$.

The only difference between a parametric model and a semi-parametric model is that in the former, the distribution of the survival time is known, and thus the model is capable of estimating risk factors in detail, whereas, in the latter, the analysis stresses the covariates rather than the risk. Moreover, the parametric model specification distinguishes itself from other models, including the non-parametric and semi-parametric models, since it can determine the distribution of survival time by employing full maximum likelihood in estimating the parameters, using residuals to measure the difference between observed and estimated survivor's time values, and give clinically useful estimates from the identified parameters [15, 16]. The failure time model with an accelerated rate is:

$$Ln(T_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \varepsilon_i, \quad i = 1, 2, ..., n \tag{2}$$

Where ε is not specified, when ε is defined, an explicitly stated accelerated failure time model becomes a parametric regression model. $T_i$represent the failure time logarithm, it is composed of $\beta_1 X_1, \beta_2 X_2, ..., \beta_p X_p$ [17, 14, 18].

3. Weibull AFT

With the Weibull regression model is currently among the most commonly used parametric models, coefficients of explanatory variables and the estimation of the baseline risk function are the outputs based on this methodology. The application is not limited to reliability engineering studies only. One broad example is the prediction of time to failure e. g., the machine's lifetime, so that optimal preventive and corrective maintenance scheduling are achieved leading to a maximum reliability of the entire plant [19]. Weibull AFT model helps in quantifying the relationship among variables and failure time of events of interest using the Weibull distribution [18]. Because it is adaptable in collecting multiple types of failure time data and can reflect a variety of hazard function shapes, the exponential decay Weibull distribution is one of the functions of survival analysis that is often used.

The shape and scale parameters keep showing the relationship at this Weibull accelerated failure time (AFT). The shape parameter assumes the constant attribute. The AFT model is the name of a transformation of response variable from logarithmic or monotone, the time of failure. This model is a type of real linear regression model [20, 21].

Here, they are the Weibull probability density function with two parameter distributions where the parameters are the scale parameter λ, and the shape parameter α defines it [22, 23]:

$$f(t;\alpha,\lambda) = \frac{(\lambda/\alpha)\,(t/\alpha)^{\lambda}\, e^{-(t/\alpha)^{\lambda}}}{(t/\alpha)} \qquad \text{where } \alpha > 0\ , \lambda > 0\ , t \geq 0 \tag{3}$$

Note that if $\alpha = 1$ we get the exponential distribution $\exp \sim (\lambda)$. In this case, the cumulative distribution function (CDF) will be [24]:

$$F(t;\alpha,\beta) = 1 - e^{\left(-t/\alpha\right)^{\lambda}} \tag{4}$$

For the Weibull distribution, the survival function is provided by [25, 26]: [26]:

$$\begin{aligned} \delta(t) &= 1 - F(t) \\ &= 1 - \left(1 - e^{(-t/\alpha)^{\lambda}}\right) = e^{(-t/\alpha)^{\lambda}} \end{aligned} \tag{5}$$

The hazard function depends on the value of α, since if $\alpha = 1$ then [19]:

$$\hbar(t) = \frac{(t)}{\delta(t)} = \lambda\alpha^{-1}\left(t/\alpha\right)^{\lambda-1} \tag{6}$$

But for different α we get different hazards for α as follows [27]:

1. $0 < \alpha < 1$ Hazard is decreasing $(1/t)$
2. $1 < \alpha < 2$ Hazard is increasing $\sqrt{t}$
3. $2 < \alpha$ Hazard is increasing t^p

The instantaneous failure rate or hazard function, rises with time if $\alpha > 1$. This is typical of a "bathtub" curve, in which failure rates are modest at first, rise over time, and then level out. A decreasing failure rate is indicated by the hazard function decreasing over time if $\alpha < 1$. The distribution becomes exponential when $\alpha = 1$, the hazard function remains constant over time [28]. The scale parameter λ affects the spread or scale of the distribution. Larger values of λ compress the distribution horizontally, leading to shorter durations. Smaller values of λ stretch the distribution, resulting in longer durations. Now suppose that:

$$\ln(T) = X\beta + \sigma\varepsilon \tag{7}$$

Here ε follows the standard Gumbel distribution$\sim G(0,1)$ and let $\lambda = e^{-X\beta}$, then Eq. (3), (4), (5) and (6), will be respectively [18]:

$$f\left(t;\sigma^{-1},e^{X'\beta}\right) = (1/\sigma)\left(e^{X'\beta}\right)^{-1}\left(t\left(e^{X'\beta}\right)^{-1}\right)^{(1/\sigma)-1}\exp\left(-\left(t\left(e^{X'\beta}\right)^{-1}\right)^{1/\sigma}\right)(1/\sigma) \tag{8}$$

$$F\left(t;\sigma^{-1},e^{X'\beta}\right) = 1 - \exp\left[-\left(t\left(e^{X'\beta}\right)^{-1}\right)^{\frac{1}{\sigma}}\right] \tag{9}$$

$$\delta\left(t;\sigma^{-1},e^{X'\beta}\right) = \exp\left[-\left(t\left(e^{X'\beta}\right)^{-1}\right)^{\frac{1}{\sigma}}\right] \tag{10}$$

$$\hbar\left(t;\sigma^{-1},e^{X'\beta}\right) = \frac{1/\sigma}{e^{X'\beta}}\left(t\left(e^{X'\beta}\right)^{-1}\right)^{\left(1/\sigma\right)-1} \tag{11}$$

The right-hand side of the equation in the Weibull AFT model represents the linear predictor, which connects the scale parameter $\lambda = e^{-X\beta}$to the variables. Generally, it is assumed that the shape parameter remains constant for every observation.

The coefficients represent the effects of the variables on the log of the survival time and, in turn, on the scale parameter of the Weibull distribution $\beta_0, \beta_1, ..., \beta_p$. By exponentiating the coefficients, you may ascertain how each covariate influences the scale parameter and, consequently, the survival time.

When the underlying distribution of survival times is predicted to follow a Weibull distribution and researchers are interested in learning how factors affect the time to an event, Weibull AFT models are frequently utilized in survival analysis. These models can show how different elements relate and how event timings accelerate or decelerate [29].

4. Sand Cat Swarm Optimization

It is one of the six species of the genus Felis. The sand cat (Felis margarita) is one of the tiniest wild cats. Sand cats are found in arid regions with stony and sandy surfaces, like the Arabian Peninsula, central Asia, and Africa. Although there isn't much of an aesthetic distinction between sand cats and domestic cats, sand cats have different

living behavior. The sand cat has more densely packed sandy to light grey fur on its hands and soles of its feet. The small, agile, and reticent sand cat exhibits distinct behavior in both hunting and dwelling [30].

A mature sand cat can weigh anywhere between one and three kilograms. The sand cat has a length of 45 to 57 centimeters on its body, with small legs next to the frontier claws, which are short and strongly curled. The tail measures between 28 and 35 centimeters, which is nearly half the length of the head and body. The rear claws have a weaker curvature and are longer. The sand cat's ears, which are 5-7 centimeters long on the sides of the head, play a significant role in its ability to forage. The fur covering on its feet protects the pads from the extreme heat and cold of the desert.

Additionally, the sand cat's fur characteristics make tracking and detection challenging. This cat is unique because it is nocturnal, underground, and secretive [31].

They use food as a source of water to ward against thirst. Most of the time, these cats lie on their backs in their burrows to dissipate body heat. Even though they face challenges in the desert, sand cats discover that the cold nighttime temperatures help them locate food. Like the Felis family, the sand cat utilizes its paws for hunting. They are skilled hunters who consume small rodents from the desert, reptiles, small birds, spiny mice, insects, and snakes [32]. The sand cats can identify subterranean rodents and insects as their prey. Sand cats have an intriguing hunting strategy; they employ their keen sense of hearing to detect low-frequency vibrations. In the outer ear, domestic cats and sand cats are the same. When it comes to the middle ear, sand cats' ear canals are longer than domestic cats', resulting in a lot of air space there.

Therefore, the smart cat, which looks for the prey and moves quickly to capture it, has the best acuity to smell the hidden game subjects, and they are also sensible for frequencies below 2 kHz [33]. When the animal picks up the speed of sound, it is used at the tackle stage of either hunting or fighting to find fish. In deer mice, each call from the partner makes all their shocking tendrils penetrate into the nervous system just to receive another palpable. Hence, being in a prime position allows the sand cat to keeplesant, where he can obtain the most prey types. Furthermore, they are outstanding in tracking both prey below the ground and the ground with the help of their huge ability to acquire food. Such a wonderful feature supports them in carrying out their tasks quickly, including detecting and apprehending any suspects. As the authors of the suggested algorithm considered wild sand cats naturally being of a solitary nature, the idea of herd behavior was taken into account in order to emphasize that the tagging procedure has implemented the concept of swarm intelligence governing this process [34].

Nature-inspired optimization algorithms (NIOA) [33, 35] are employed in wrapper approaches as search strategies to find informative features that have been applied to address a variety of optimization issues, including technical issues [36, 37], feature selection [38, 39] and power transformer fault diagnosis [40].

Figure 1. Sand Cat Swarm in Nature

4.1. Model and Methods

Since there are Nsand cat individuals (solutions) in the population of SCSO, each of which has D dimensions, the population vector in SCSO comprises a $N \times D$ dimensional matrix. For every sand cat, the answer is represented by $X_i = (x_{i1}, x_{i2}, ..., x_{id})$. The SCSO algorithm's hearing ability is advantageous when detecting sand cats at low frequencies. Therefore, it is believed that the sand cat's sensitivity range begins at 2 kHz and ends at 0 in order to look for prey. Let's represent the sensitivity level in SCSO as follows:

$$r_g = \delta_M - \left(\frac{\delta_m \times t}{T}\right) \tag{12}$$

Suppose that $\delta_m = 2$, as it is derived from the sand cats' auditory features, t is the current iteration, and T is the upper limit of repetitions. The trade-off between the exploration and exploitation phases is established by the R parameter, which is calculated as follows:

$$R = ((r_g \times 2) \times rand(0,1)) - r_g \tag{13}$$

Using $rand(0,1)$ to get a random number between 0 and 1. The coordinates of the sand cat in the search space at iteration are denoted by the position vector of that sand cat: The potential of a given solution to the customer across the sensitivity spectral range is as defined by r value and formula:

$$r = r_g \times rand(0,1) \tag{14}$$

The value of $R \in [1,-1]$, determines where the sand cat will go next. The SCSO method is based on the procedure of using stimulation and leading it to chase after its target when $|R|$ is less than or equal to 1. The mathematical expression for prey exploitation in SCSO is as follows:

$$\mathrm{X}_{rand} = |rand(0,1) \times X_{best} - X(t)| \tag{15}$$

Where X_{rand} determines the separation in the associated iteration t between the current position $X(t)$ and the best position X_{best}. For the associated search agent X, the location update is shown in X_{t+1} as follows:

$$X_{t+1} = X_{best} - rand(0,1) \times X_{rand} \times \cos(\theta) \tag{16}$$

There is a fact that discernment of the sand cats cannot be done without the special micro circles. A random angle θ selected from a roulette wheel determines (between 0 and 360°(the direction of each movement. Consequently, the mathematical representation of pursuing prey (exploration) is [41]:

$$cp = floor(N \times rand(0,1) + 1) \tag{17}$$

$$X_{\mathrm{Candidate}}(t) = X(cp,:) \tag{18}$$

$$X_{t+1} = r \times (X_{\mathrm{Candidate}}(t) - rand(0,1) \times X(t)) \tag{19}$$

$X_{\mathrm{Candidate}}(t)$denotes a candidate position that is randomly selected.
The pseudo-code of SCSO Algorithm is [39]:

1. N=number of population and T=Max number of iterations
2. Establish the sand cat population X_i
3. **While** $t \leq T$**do**
4. Determine the fitness of each sand cat by evaluating the objective function
5. Determine X_{best}
6. Determine $rg$when $s_M = 2$
7. **For** $i = 1$ **to** N **do**
8. Calculate R and r
9. **For** $j = 1$ **to** D **do**
10. Using the Roulette wheel selection method, a value $0 \leq \theta \leq 360$ is chosen randomly
11. **if** $((-1 <= R)\,\&\&\,(R <= 1))$ **then**
12. $X_{rand} = \left|rand(0,1) \times X_{best,j} - X_{(i,j)}\right|$
13. $X_{(i,j)} = X_{best,j} - rand(0,1) * X_{rand} * COS(\theta)$ // update position using (16)
14. **else**
15. $cp = floor(N * rand(0,1) + 1)$
16. $X_{candidate} = X(cp,:)$
17. $X_{(i,j)} = r \times \left(X_{candidate,j} - rand(0,1) \times X_{(i,j)}\right)$ //update position using (19)
18. **End if**

19. **End for**
20. **End for**
21. $\mathrm{t} = \mathrm{t} + 1$
22. **End While**

Return X_{best}

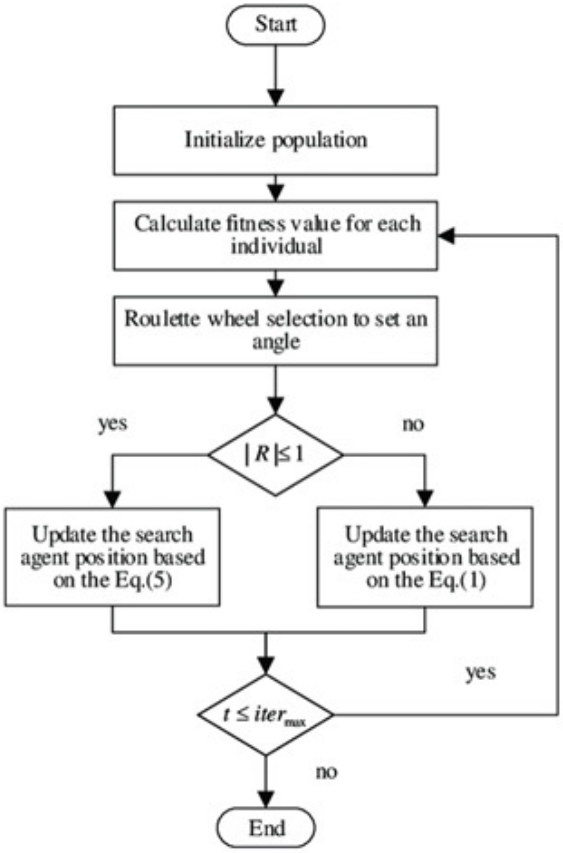

Figure 2. The flowchart of SCSO

Table 1. A simple example

x_1	x_2	x_3	x_{p-1}	x_p
0	1	0	0	1

5. The Proposed approach

Because of their many useful characteristics, including ergodicity, mixing property, and sensitivity to beginning circumstances, nonlinear chaotic maps in chaotic systems are significant in the fields of engineering, biology, and economics. In this paper, we use SCSA to carry out the feature selection. An optimization problem with a search on interval [0,1]. In order to enhance SCSA performance, chaotic maps are taken into consideration in this paper. It is believed that chaotic maps allow SCSA to be freed from the trapping of local optima, which would help

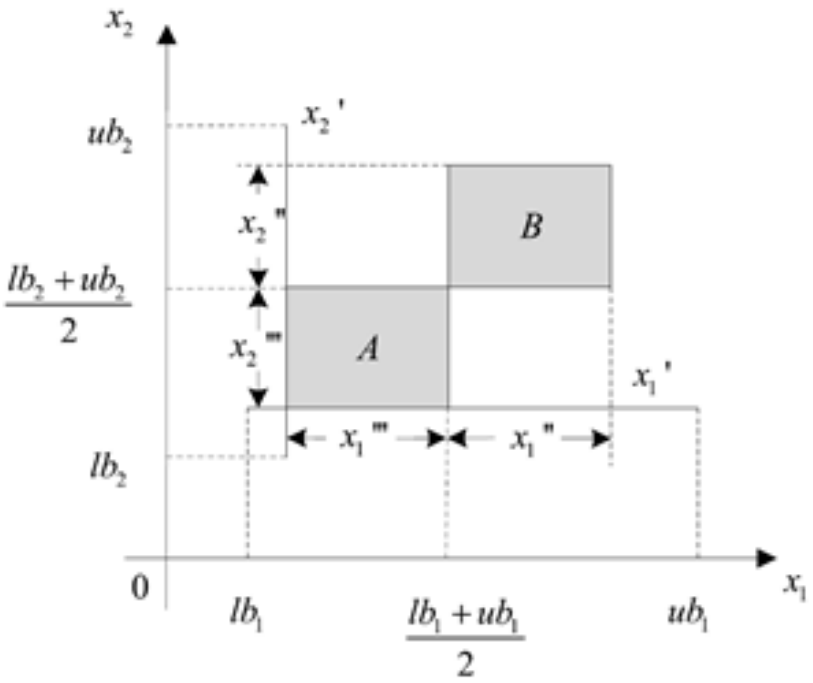

Figure 3. The arrangement of simulated opposing and reflecting points inside the search area

achieve a faster convergence of the selection of variables in the Weibull model. This paper uses ten chaotic maps, the random parameter values of SCSA are manipulated employing these chaotic maps. Due to the potential impact of initial values on the fluctuation pattern, we have standardized the initial point for all chaotic maps to 0.7, while keeping the remaining parameters unchanged. Any of the features could be declared as binary decision variables that indicate the degree to which a particular feature is relevant within the context of the model [42]. For the purpose of the current consideration, let it be defined that we have a vector consisting of D elements, which would be the complete set of features, DIf using the vector model, any component of a vector represents a feature, and if selecting the feature, it equals '1' whereas if non-selected, it equals '0'. Therefore, the SCSO approach will be more suitable for the continuous space, which characterizes the definition of feature selection as an optimization problem rather than the discrete space. Realizing this, we must stipulate that S is discrete. Therefore, the configuration of our suggested method is as follows:

Step (1): The number of cats is $\eta_c = 40$, $\varphi_0 = 1$ "where $\varphi_0$represents the attractiveness at 0 "$\alpha = 0.1$", α is the randomization parameter" and the upper limit of iterations is $t_{\max} = 200$.

Step (2): The cats' positions in the original SCSO are produced following a continuous uniform distribution on the interval [0,1]. The proposed chaotic maps utilize the maps outlined in Table 1.

Step (3): The fitness function is formally defined as:

$$fitness = \min\left[1/n\left(\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right)\right] \tag{20}$$

Step (4): update of the position of cats Coordinates depends on Eq (16).

Step (5): Steps 3 and 4 are iterated until a $t_{\max}$ is reached.

Table 2. The Description of Chaotic Maps

Name	Definition	Range
Chebyshev	$x_{K+1} = \cos\left(\frac{1}{K}\cos^{-1}(x_K)\right)$	(-1,1)
Circle	$x_{K+1} = \mathrm{mod}\left(x_K + 0.2 - \frac{0.5}{2\pi}\sin(2\pi x_K), 1\right)$	(0,1)
Guass/mouse	$x_{K+1} = \begin{cases} 1 & x_K = 0 \\ 1/\mathrm{mod}(x_K, 1) & otherwise \end{cases}$	(0,1)
Iterative	$x_{K+1} = \sin((0.7)\pi/x_K)$	(-1,1)
Logistic	$x_{K+1} = (1 - x_K)4x_K$	(0,1)
Piecewise	$x_{K+1} = \begin{cases} \frac{x_K}{0.4} & 0 \le x_K < 0.4 \\ \frac{x_K - 0.4}{0.1} & 0.4 \le x_K < 0.5 \\ \frac{0.6 - x_K}{0.1} & 0.5 \le x_K < 0.6 \\ \frac{1 - x_K}{0.4} & 0.6 \le x_K < 1 \end{cases}$	(0,1)
Sine	$x_{K+1} = \sin(\pi x_K)$	(0,1)
Singer	$x_{K+1} = 1.07\left(7.86x_K - 23.31(x_K)^2\right) + 28.8(x_K)^3 - 13.30288(x_K)^4$	(0,1)
Sinusoidal	$x_{K+1} = 2.3x_K\sin(\pi x_K)$	(0,1)
Tent	$x_{K+1} = \begin{cases} \frac{x_K}{0.7} & x_K < 0.7 \\ \frac{10}{3}(1 - x_K) & x_K \ge 0.7 \end{cases}$	(0,1)

6. Simulation Results

We used various regularization techniques in conjunction with the AFT approach to conduct the simulation studies (SCSO, Elastic net, Lasso, L1/2, and MCP) [43] to assess their predictions' accuracy. Bender's work served as the model for the AFT model simulation schemes. The following is how our simulation data were produced [44]:
Step (1): Calculate the correlation coefficient$\rho$and build an array $\delta_0, \delta_{i1}, \delta_{i2}, ..., \delta_{ip}$ where $i = 1, 2, ...,n$ are independent following standard normal distribution, and set:

$$x_{ij} = \delta_{ij}\sqrt{1-\rho} + \delta_{i0}\sqrt{\rho}$$

Where $j = 1, 2, ...,p$.
Step (2): The independent Survival time $y_i$as follow:

$$y_i = \exp\left(\sum_{j=1}^{p}\beta_{ij}x_{ij}\right)$$

Step (3): Let y'_i where $i = 1, 2, ...,m$and $m$is the number of the censored sample follow a random distribution. The number of censored data is determined by the censoring rate.
Step (4): suppose that $y_i = \min(y_i, y'_i)$, Therefore, the values of y_i, δ_i (where δ_i is the survival function of the Kaplan-Meier estimator), and x_i are considered as the observed data and are utilized in the AFT model.

Since our simulation depends on Leukemia cancer data, we establish the size or scale of the predictor genes with $p = 1000$, and there are eight non-zero of the remaining coefficients $\beta_1 = 1.5$, $\beta_2 = 1.1$, $\beta_3 = 0.81$, $\beta_5 = 3.5$, $\beta_9 = 4.6$, $\beta_{10} = -1.7$, $\beta_{11} = 0.5$and $\beta_{13} = 1.3$, the remaining coefficients 992 genes are equal to zeros. Note that x_1, x_2x_3, x_5, x_9, x_{10}, x_{11} and x_{13} are important variables. The right censored (C) value is set at 10%, 20% and 40%, the training sample size is determined by three conditions $n = (100, 300, 500)$, with correlation coefficient$\rho = 0.4$.As a technique to optimize fidelity, a model supported by different regularization regulations can be tested on 50 data sets for a training set and then to obtain a prediction on a testing set consisting of 50 data. For instance, each outcome of those operations was computed based on the 200 out-turn after opening 200 doors.

The data represents the mean performance for the Total # of features and the Correct # of features chosen by each regularization approach. These 200 iterations indicate the performance accomplished in 200 repeated experiments.

Outcomes average, to be precise, the total of selected features and the number of correct features are shown in Table (1) for the candidates of the beach regularization approach alongside 200 repeated tests.

When the training sample size is extremely limited (n=50), all of the techniques proved to be quite challenging in accurately identifying the appropriate genes. As the value of n increases, the ability to identify more accurate

non-zero features also increases. When =0, all the approaches exhibit superior performance compared to when =0.3.

An assessment of the results achieved by the SCSO maps is made. Furthermore, the proposed system has all its performance measures evaluated against that of the existing SCSO. In this case, the performance quality is evaluated with mean-squared error (MSE) on both the training and testing databases, as well as the number of chosen variables. The identified themes' details are presented in tables 2 & 3 below.

Table 3. The performance of the used methods for the train data when $n = 100$

Method	**No. of selected variable**	**MSE**	**No. of selected variable**	**MSE**	**No. of selected variable**	**MSE**
	C=10%		C=20%		C=40%	
SCSO	27	4.658	27	5.858	27	7.558
Chebyshev	23	4.381	23	5.611	23	7.311
Circle	20	4.124	20	5.224	20	7.124
Guass/mouse	22	4.551	22	5.821	22	7.601
Iterative	25	4.381	25	5.631	25	7.331
Logistic	26	4.711	26	6.011	26	7.812
Piecewise	19	3.864	19	4.994	19	6.694
Sine	17	3.646	17	4.946	17	6.646
Singer	13	3.442	13	4.142	13	5.842
Sinusoidal	14	3.580	14	4.625	14	6.325
Tent	11	3.124	11	4.324	11	6.024

Table 4. The performance of the used methods for the train data when $n = 300$

Method	**No. of selected variable**	**MSE**	**No. of selected variable**	**MSE**	**No. of selected variable**	**MSE**
	C=10%		C=20%		C=40%	
SCSO	27	3.558	27	4.548	27	6.248
Chebyshev	23	3.321	23	4.221	23	6.221
Circle	20	3.014	20	4.114	20	6.114
Guass/mouse	22	3.141	22	4.041	22	5.741
Iterative	25	3.231	25	4.131	25	5.331
Logistic	26	3.591	26	4.613	26	6.313
Piecewise	19	2.824	19	3.694	19	5.394
Sine	17	2.436	17	3.336	17	5.036
Singer	13	2.342	13	3.242	13	5.042
Sinusoidal	14	2.488	14	3.299	14	5.147
Tent	11	2.124	11	3.024	11	4.724

Furthermore, it is evident that the Tent map achieves the lowest Mean Squared Error (MSE) compared to the other chaotic maps utilized. Compared to the SCSO, the Tent map exhibited a decrease of approximately 32.932%-20.291%, 40.303%-24.392% and 58.34%-47.63% from tables 1,2 and 3, respectively.

7. Real application

To evaluate the efficiency of the proposed strategy, using the genes dataset of a real-world case study is recommended. Table 6 condenses the entire content, datasets, and their justification for this study. First and foremost, the data set used is the Diffuse large B-cell lymphoma dataset (DLBCL) [45]. There are a total of 240 samples from people with lymphoma. Each patient's data comprises 7399 gene expression measures and their corresponding survival time, whether censored or not. The second data collection is the Dutch breast cancer dataset (DBC), which has 79 instances with 30 features. It is a dataset of the 295 treatments administered to breast cancer patients. The molecular profile of every patient has 4919 gene expression readings. They also generated RNA-Seq

Table 5. The performance of the used methods for the train data when $n = 500$

Method	No. of selected variable	MSE	No. of selected variable	MSE	No. of selected variable	MSE
	C=10%		C=20%		C=40%	
SCSO	27	2.458	27	3.158	27	4.055
Chebyshev	23	2.221	23	3.121	23	3.321
Circle	20	2.014	20	2.714	20	3.114
Guass/mouse	22	2.141	22	3.241	22	3.241
Iterative	25	2.231	25	2.931	25	3.331
Logistic	26	2.511	26	3.211	26	3.711
Piecewise	19	1.724	19	2.424	19	4.114
Sine	17	1.336	17	2.036	17	2.436
Singer	13	1.242	13	2.042	13	2.342
Sinusoidal	14	1.438	14	2.138	14	2.498
Tent	11	1.024	11	1.724	11	2.124

profiles of the patient's tumors [46]. The third class is cancer of the lung (LC). The data is a record of 86 patients who have been diagnosed with lung cancer. The expression level data of a particular gene is 7129, and it also covers survival time, such as alive or censored [47].

Table 6. The specifics of the three utilized authentic microarray datasets

Dataset	Sample	Gene	Censored
DLBCL	240	7399	102
DBC	295	4919	207
LC	86	7129	62

Akaaki's criterion (AIC) and Bayes' criterion (BIC) [48] will be used as methods for selecting variables according to formulas [49]:

$$AIC = 2k - 2\log(L)$$

$$BIC = 2Log(L) - k \times Log(n)$$

Where $k$equal to the number of explanatory variables.

By comparing the proposed model to AIC and BIC, we shall exhibit efficiency. The step to execute this is to divide the expressions of genes in each dataset into a training and an unseen dataset by random and selecting 70% of the data for the training dataset and the rest 30% for testing dataset. The evaluation of our algorithm for predicting the results is done by using time-dependent receiver-operator characteristics curves (ROC) for the censored data and area under the curve (AUC) A Table 6 summarizes the fact that world application is the main pillar in the real-life results.

Table 6 contains the average scores of the three real-world datasets used and the methods used to implement them. This evidence is enough to infer that, given that AIC selects many more genes than the BIC and the SCSO, there already exist huge differences between the three approaches. From the 3 applications, the SCSO then chose the fewest genes in the last subset.

This provides a reference for the model performance evaluation where the mean AUC of the two kinds of data is provided to compare and show the relationship of both training and testing datasets. Tables 3 and 4 demonstrate their embodiment of the objective (standpoint for interval measurements) and a subjective (perceived distance) distance. The results demonstrated that the Sine map, with accuracy levels of 97.7% for DLBCL, 98.6% for DRBC, and 98.9% for LC datasets, respectively, achieved the highest precision; the map Singer came in second place in terms of its efficiency with a percentage 95.2% for DLBCL, 97.9% for DRBC and 98.2% for LC, Eventually, the tool to quantify the performance was settled on by an AUC metric (the result optimal of the tool with a crucial purpose). Accordingly, it could be safely concluded that the supremacy of the multilayer perceptron is derived from the fact that the AIC and BIC were calculated in nearly the same ways across the datasets, which did not show different performances. Then SCSO's I-HO "contribution" is expected to be more superb than AIC's.

Table 7. The outcomes of the chosen genes

	DLBCL	**DBC**	**LC**
AIC	98	75	82
BIC	77	59	73
SCSO	31	27	21
Chebyshev	25	23	21
Circle	22	20	18
Guass	23	22	19
Iterative	26	25	19
Logistic	27	26	24
Piecewise	20	18	16
Tent	18	15	13
Singer	14	13	11
Sinusoidal	15	13	12
Sine	12	9	8

Table 8. The outcomes of the chosen genes

	DLBCL	**DBC**	**LC**
AIC	0.874	0.879	0.881
BIC	0.881	0.895	0.961
SCSO	0.901	0.923	0.940
Logistic	0.908	0.928	0.951
Iterative	0.912	0.932	0.953
Chebyshev	0.918	0.939	0.958
Guass	0.927	0.947	0.967
Circle	0.929	0.951	0.969
Piecewise	0.934	0.956	0.971
Tent	0.937	0.962	0.973
Sinusoidal	0.944	0.967	0.978
Singer	0.952	0.979	0.982
Sine	0.977	0.986	0.989

Table 9 presents the results of the test data set to prove the superiority of the proposed method over the remaining methods in terms of AUC. The Sine is a network with the biggest AUC at 93.2%, 96.2% and 98.9% in DLBCL, DBC and LC datasets, respectively.

Table 9. The AUC results for the testing dataset

	DLBCL	**DBC**	**LC**
AIC	0.736	0.785	0.791
BIC	0.759	0.801	0.871
SCSO	0.776	0.829	0.851
Logistic	0.784	0.832	0.868
Iterative	0.787	0.838	0.869
Chebyshev	0.793	0.851	0.871
Guass	0.802	0.853	0.877
Circle	0.808	0.856	0.879
Piecewise	0.819	0.867	0.881
Tent	0.826	0.878	0.885
Sinusoidal	0.879	0.882	0.910
Singer	0.897	0.885	0.942
Sine	0.932	0.962	0.989

This was proof that the Sine, Singer, Sinusoidal, Tent, Piecewise, Circle, Guass, Chebyshev, Iterative and Logistic maps was highly accurate, respectively, in identifying the ones who were actually at risk of having cancer with a greater than 0.95 probability.

For more comparison. Our proposed method, ”Sine”, was compared with other classic feature selection methods, such as the Least Absolute Shrinkage and Selection Operator (LASSO) and Adaptive Least Absolute Shrinkage,

additionally other metaheuristic algorithms such as Particle Swarm Optimization (PSO) and Genetic Algorithms (GA); these comparisons are explained in Table 10 as follows:

Table 10. The AUC results for the testing dataset

	DLBCL	**DBC**	**LC**
Sine	0.932	0.962	0.989
LASSO	0.605	0.697	0.735
ALASSO	0.718	0.789	0.841
PSO	0.817	0.849	0.905
GA	0.894	0.907	0.945

From Table (10) above, we note that the proposed algorithm (Sine) strongly outperformed the classical methods and was better than the PSO algorithm and the GA algorithm, which came in second place according to the AUC criterion, followed by the PSO algorithm.

8. Conclusion

Selection variables influence quality biases in Weibull AFT compliance models. For feature selection, this paper employed a swarm of Sand cat optimization algorithms across ten architectural styles as a possible alternative. Simulations and real data visually illustrated the superior performance of the proposed algorithms in terms of MSE on training data. Furthermore, it highlighted that the importance in this scenario exceeded all other factors.

REFERENCES

1. T. G. Clark, M. J. Bradburn, S. B. Love, and D. G. Altman, *Survival analysis part I: basic concepts and first analyses*, British journal of cancer, vol. 89, no. 2, pp. 232-238, 2003.
2. S. P. Khanal, V. Sreenivas, and S. K. Acharya, *Accelerated failure time models: an application in the survival of acute liver failure patients in India*, Int J Sci Res, vol. 3, no. 6, pp. 161-6, 2014.
3. M. K. Hassan and H. T. Abed, *Comparison of survival models to study determinants liver cancer*, journal of Economics And Administrative Sciences, vol. 27, no. 125, 2021.
4. P. Lambert, D. Collett, A. Kimber, and R. Johnson, *Parametric accelerated failure time models with random effects and an application to kidney transplant survival*, Journal of Economics and Administrative Sciences,vol. 28, no. 133, pp. 82-96,2022.
5. J. D. Kalbfleisch and R. L. Prentice, *Estimation of the average hazard ratio*, Biometrika, vol. 68, no. 1, pp. 105-112, 1981.
6. B. S. Kim and G. Maddala, *Estimation and specification analysis of models of dividend behavior based on censored panel data*, in Panel Data Analysis, 1992: Springer, pp. 111-124.
7. S. S. Kumar and R. Elangovan, *VARIABLE SELECTION FOR THE ACCELERATED FAILURE TIME MODELS*, Advance and Innovative Research, p. 79, 2018.
8. L. A. Gelfand, D. P. MacKinnon, R. J. DeRubeis, and A. N. Baraldi, *Mediation Analysis with Survival Outcomes: Accelerated Failure Time vs. Proportional Hazards Models*, Front Psychol, vol. 7, p. 423, 2016, doi: 10.3389/fpsyg.2016.00423.
9. J. Abolghasemi, M. N. Toosi, M. Rasouli, and H. Taslimi *Survival analysis of liver cirrhosis patients after transplantation using accelerated failure time models*, Biomedical Research and Therapy, vol. 5, no. 11, pp. 2789-2796, 2018.
10. J. L. Moran et al., *Modelling survival in acute severe illness: Cox versus accelerated failure time models*, Journal of Evaluation in Clinical Practice, vol. 14, no. 1, pp. 83-93, 2008.
11. C. R. Andersen, J. Wolf, K. Jennings, D. S. Prough, and B. E. Hawkins, *Accelerated failure time survival model to analyze Morris water maze latency data*, Journal of neurotrauma, vol. 38, no. 4, pp. 435-445, 2021.
12. W. Weibull, *A statistical distribution function of wide applicability*, Journal of applied mechanics, 1951.
13. A. Karimi, A. Delpisheh, and K. Sayehmiri, *Application of accelerated failure time models for breast cancer patients' survival in Kurdistan Province of Iran*, Journal of Cancer Research and Therapeutics, vol. 12, no. 3, pp. 1184-1188, 2016.
14. H. Chai, Y. Liang, and X.-Y. Liu, *The L1/2 regularization approach for survival analysis in the accelerated failure time model*, Computers in biology and medicine, vol. 64, pp. 283-290, 2015.
15. D. W. Hosmer and S. Lemeshow, *Survival analysis: applications to ophthalmic research*, American journal of ophthalmology, vol. 147, no. 6, pp. 957-958, 2009.
16. H. Wang, H. Dai, and B. Fu, *Accelerated failure time models for censored survival data under referral bias*, Biostatistics, vol. 14, no. 2, pp. 313-326, 2013.
17. J. Huang, S. Ma, and H. Xie, *Regularized estimation in the accelerated failure time model with high-dimensional covariates*, Biometrics, vol. 62, no. 3, pp. 813-820, 2006.
18. E. Liu, R. Y. Liu, and K. Lim, *Using the Weibull accelerated failure time regression model to predict time to health events*, Applied Sciences, vol. 13, no. 24, p. 13041, 2023.

19. H. Rasheed, T. Al-Baldawi, and N. Al-Obedy, *COMPARISON OF SOME BAYES'ESTIMATORS FOR THE WEIBULL RELIABILITY FUNCTION UNDER DIFFERENT LOSS FUNCTIONS*, Iraqi Journal of Science, vol. 53, no. 2, pp. 362-366, 2012.
20. S. A. Kareem, *A Comparison Between Two Shape Parameters Estimators for (Burr-XII) Distribution*, Baghdad Science Journal, vol. 17, no. 3 (Suppl.), pp. 0973-0973, 2020.
21. J. J. Gorgoso-Varela and A. Rojo-Alboreca, *Use of Gumbel and Weibull functions to model extreme values of diameter distributions in forest stands*, Annals of forest science, vol. 71, no. 7, pp. 741-750, 2014.
22. R. A. Abdulwahab, A. H. Shaban, and E. H. Nasser, *Characteristics of Electrical Power Generation by Wind for Al-Tweitha Location Using Weibull Distribution Function*, Iraqi Journal of Science, pp. 1120-1126, 2014.
23. I. Alkanani and D. Majeed, *Parameters Estimation for Modified Weibull Distribution Based on Type One Censored Samplest*, Iraqi Journal of Science, vol. 54, no. Mathematics conf, pp. 822-827, 2013.
24. A. ALNaqeeb and A. Hamad, *Suggested method of Estimation for the Two Parameters of Weibull Distribution Using Simulation Technique*, Iraqi Journal of Science, vol. 54, no. Mathematics conf, pp. 744-748, 2013.
25. N. Juhan, N. Abd Razak, Y. Z. Zubairi, N. N. Naing, C. Hussin, and M. A. Daud, *Comparison of stratified Weibull model and Weibull accelerated failure time (aft) model in the analysis of cervical cancer survival*, Jurnal Teknologi, vol. 78, no. 6-4, pp. 21-26, 2016.
26. M. R. Karim and M. A. Islam, *Reliability and survival analysis*, Springer, 2019.
27. S. M. Jaafar, B. I. Abdulwahhab, and I. A. Hasson *Best estimation for the Reliability of 2-parameter Weibull Distribution*, Baghdad Science Journal, vol. 6, no. 4, pp. 705-710, 2009.
28. G. K. David and K. Mitchel, *Survival analysis: a Self-Learning text*, Spinger, 2012.
29. R. Elangovan, R. Mohanasundari, and E. Susiganeshkumar, *ACCELERATED FAILURE TIME MODEL FOR SURVIVAL ANALYSIS*.
30. F. R. Cole and D. E. Wilson, *Felis margarita (Carnivora: Felidae)*, Mammalian Species, vol. 47, no. 924, pp. 63-77, 2015.
31. G. Huang, J. Rosowski, M. Ravicz, and W. Peake, *Mammalian ear specializations in arid habitats: structural and functional evidence from sand cat (Felis margarita)*, Journal of Comparative Physiology A, vol. 188, pp. 663-681, 2002.
32. M. Abbadi, *Radiotelemetric observations on sand cats (Felis margarita) in the Arava Valley*, Israel Journal of Zoology, vol. 36, pp. 155-156, 1989.
33. A. Seyyedabbasi and F. Kiani, *Sand Cat swarm optimization: A nature-inspired algorithm to solve global optimization problems*, Engineering with Computers, vol. 39, no. 4, pp. 2627-2651, 2023.
34. D. Wu, H. Rao, C. Wen, H. Jia, Q. Liu, and L. Abualigah, *Modified sand cat swarm optimization algorithm for solving constrained engineering optimization problems*, Mathematics, vol. 10, no. 22, p. 4350, 2022.
35. E. Pashaei, *An Efficient Binary Sand Cat Swarm Optimization for Feature Selection in High-Dimensional Biomedical Data*, Bioengineering, vol. 10, no. 10, p. 1123, 2023.
36. F. Kiani, F. A. Anka, and F. Erenel, *PSCSO: Enhanced sand cat swarm optimization inspired by the political system to solve complex problems*, Advances in Engineering Software, vol. 178, p. 103423, 2023.
37. Y. Yu, Y. Li, J. Li, X. Gu, and S. Royel, *Nonlinear characterization of the MRE isolator using binary-coded discrete CSO and ELM*, International Journal of Structural Stability and Dynamics, vol. 18, no. 08, p. 1840007, 2018.
38. A. Qtaish, D. Albashish, M. Braik, M. T. Alshammari, A. Alreshidi, and E. J. Alreshidi, *Memory-based Sand Cat Swarm Optimization for Feature Selection in Medical Diagnosis*, Electronics, vol. 12, no. 9, p. 2042, 2023.
39. A. Seyyedabbasi, *Binary sand cat swarm optimization algorithm for wrapper feature selection on biological data*, Biomimetics, vol. 8, no. 3, p. 310, 2023.
40. W. Lu, C. Shi, H. Fu, and Y. Xu, *A power transformer fault diagnosis method based on improved sand cat swarm optimization algorithm and bidirectional gated recurrent unit*, Electronics, vol. 12, no. 3, p. 672, 2023.
41. Y. Li and G. Wang *Sand cat swarm optimization based on stochastic variation with elite collaboration*, IEEE Access, vol. 10, pp. 89989-90003, 2022.
42. R. A. R. Mahmood, A. Abdi, and M. Hussin, *Performance evaluation of intrusion detection system using selected features and machine learning classifiers*, Baghdad Science Journal, vol. 18, no. 2 (Suppl.), pp. 0884-0884, 2021.
43. A. N. Alkhateeb and Z. Y. Algamal, *Variable selection in gamma regression model using chaotic firefly algorithm with application in chemometrics*, Electronic Journal of Applied Statistical Analysis, vol. 14, no. 1, pp. 266-276, 2021.
44. G. I. Sayed, A. Darwish, and A. E. Hassanien, *A new chaotic whale optimization algorithm for features selection*, Journal of classification, vol. 35, pp. 300-344, 2018.
45. A. Rosenwald et al., *The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma*, New England Journal of Medicine, vol. 346, no. 25, pp. 1937-1947, 2002.
46. H. C. Van Houwelingen, T. Bruinsma, A. A. Hart, L. J. Van't Veer, and L. F. Wessels, *Cross-validated Cox regression on microarray gene expression data*, Statistics in medicine, vol. 25, no. 18, pp. 3201-3216, 2006.
47. D. G. Beer et al., *Gene-expression profiles predict survival of patients with lung adenocarcinoma*, Nature medicine, vol. 8, no. 8, pp. 816-824, 2002.
48. M. K. Awad and H. A. Rasheed, *Bayesian Estimation for the Parameters and Reliability Function of Basic Gompertz Distribution under Squared Log Error Loss Function*, Iraqi Journal of Science, pp. 1433-1439, 2020.
49. S. I. Vrieze, *Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC)*, Psychological methods, vol. 17, no. 2, p. 228, 2012.