IRAQI STATISTICIANS JOURNAL

https://isj.edu.iq/index.php/isj

ISSN: 3007-1658 (Online)

Variable Selection in Weibull Accelerated Survival Model Based on lemur's Optimization Algorithm

Ahmed Naziyah Abdullah[1], Qutaiba Nabil Nayef AL-Qazaz[2]

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

ARTICLE INFO

Article history:
Received 27/11/2024
Revised 27/11/2024
Accepted 14/1/2025
Available online 15/5/2025

Keywords:
Weibull Distribution
High Dimensional
Accelerated Failure Time
Feature Selection
Lemur's Algorithm

ABSTRACT

Survival models play a key role in analyzing time-to-event data. Accelerated time models are useful when the effect of the covariates on the survival time is proportional throughout the follow-up period. The well-known Weibull accelerated failure-time model (AFT), as an accelerated time model, is widely employed in survival analyses. The high-dimensional variables picked, considering most of the low-dimensional AFT model-based variable selectors, assume that the covariate effects are constant throughout the period. However, the Weibull AFT model includes time transformations enabling constant, increasing, or decreasing impacts during the entire study, and thus, it is a more flexible approach. Lemurs' optimization algorithm (LOA), a new powerful algorithm, is employed in the covariate selection. It is important to select appropriate significant covariates for AFT models in practice. However, like all other high-dimensional data, the lemur faces the curse of dimensionality problem in Weibull AFT model-based variable selection. Thus, it is necessary to have a purposeful variable selection algorithm for Weibull AFT models that considers the increasing, constant, or decreasing effects of covariates.

1. Introduction

Information sets that quantify the time for an event to occur is called the survival data. Other variables of interest include the employment duration of heart transplant recipients and their survival rate. Some factors must be taken into account when doing any analysis involving such data [1, 2].

The survivor time variable becomes an event time variable indicating how long it will take for a particular event to occur, along with some set of such hypothesised independent variables involving survival data. The concurrent variables, also called independent variables, also may be continuous, for example age or temperature, or discrete such as sex or race. In most medical data, the system of the event of interest can be biological or physical, which is what are the engineering data [3, 4]. The two main objectives in survival analysis include modelling the underlying distribution of the failure time variable and determining how it depends on the independent variables.

In both cases the only rate of failure that is clearly seen to be related to a given censored observation is the lower boundary. But it proposed that such observation is right censored. The study is extended with an additional variable to detect which failure times are censored and which are observed event timings. The failure time may be outside a given interval (interval censored) or smaller than a specified value (left censored) more broadly; that is, the time failure only is known to be between a certain interval even if not

* Corresponding author.E-**mail** address: ahmed.alkhateeb@uomosul.edu.iq
https://doi.org/10.62933/0rq0gs73

within a given interval. Several types of potential censorship strategies arise in survival analysis. Some comparable censoring circumstances are reviewed in [5] and several filtering systems are discussed in [6]. Data containing censored observations cannot be analysed, ignoring the censored observations, as censored observations are often from longer lived persons, and thus right censored, among other reasons. We must analyse with censoring in mind and correctly use the censored and uncensored observations. AFT survival models can be applied on the Acute liver failure patients in India, if a correct survival time distribution is selected [2].

AFT is a completely parametric model and as such is permitted to reach conclusions that would be intractable in a non, or semi-parametric framework, such as estimating tail probabilities. AFT models are linear mixed models with the survival data log transformed, with censoring and correlation accounted for in the survival data. In the compromise, one must accept a specific distribution of survival times, that could be revised. The AFT model specifies that the effect of a covariate will speed up or slow down a disease's course by some fixed amount. In AFT models, the covariate affects the distribution of the response variable and of the time scale in full. Furthermore in the Proportional Hazard (PH) models the effect of covariate is multiplicative in terms of hazard rates [7]. It is well known that parametric survival models do not rely on restrictive assumption of PH. Interpretability, managing censored data, and versatility in mediation analysis with survival outcomes make accelerated failure time (AFT) models preferred to proportional hazards (PH) models.

Recognizing these distinctions is important, and researchers should select a suitable model for their mediation analysis [4, 7]. Specific assumptions are made as a basis for these basic conditional models of Cox type and AFT models and the connexion between the survival and coincidence of the variables. If you assume that you meet another common, and rather flexible model of Cox, which assumes that of the proportionality of hazards, you have another common model. In the case of the Cox model, the PH assumption requires all independent variables to remain independent of time in the final model and, in particular, the risk ratio of the event need not change over a specific period. This premise allows us to extend further investigation of the output classification process which will be more tractable than the parametric model interpretation [8]. Distribution of survival times can be analysed with AFT models, and these are particularly helpful if the proportional hazards assumption does not hold [9]. Furthermore, Parametric survival models provide several utilities for parameter estimation based on survival instead of survival hazard, and use of the complete likelihood for parameter estimation [2]. Considering that AFT models are used [10] as the suggested analytical tool to examine latency to the platform in Multi-Walled Carbon Nanotubes on research Morris water maze, AFT, logistic regression model and analysis of variance (ANOVA) is compared.

One of the most obvious [11] probabilistic distributions capable of studying survival data is the Weibull distribution, which has its application started within the industrial field in order to perform reliability testing. Just as linear modelling's normal distribution is crucial to the parametric analysis of survival data this distribution is just as necessary. Although the AFT can be used to compare survival times, the assumption of the PH is used to compare hazards [12].

Feature selection has been widely applied in various research fields. With its application in more and more areas, researchers have proposed algorithms to improve the accuracy of feature selection and find the globally optimal feature combination. However, most of the proposed methods mainly focus on categorizing the data and solving different patterns. Experimental results have shown that the existing methods are time-consuming, and the accuracy of feature selection is far from the global optimal solution. They are often trapped in a locally optimal situation, with a greater probability of falling into a local maximum and greater computational errors. Therefore, it is essential

to construct a new, high-precision feature selection method.

Feature selection, also known as variable selection, is an important part of constructing an effective model in machine learning. It removes unimportant or redundant features and keeps the features that are important and helpful for training the model. Having fewer features can reduce the computation burden of model training and improve the classification accuracy.

To solve the problems mentioned above, this paper proposes an improved Lemurs optimization algorithm to strengthen the search for the best features. The improvements of the new algorithm are two-fold: Firstly, Lévy Flight (It is a mathematical model used to describe movement in living organisms characterized by long jumps or short steps to balance exploration and exploitation) is introduced into the search process to escape from local minima, thus enabling a more global exploration for the best features. Secondly, a so-called chaotic Map is used to speed up the algorithm and achieve better solutions faster.

2. Regression Models for Survival Data

We require other forms of regression models because linear regression model is not a feasible solution given the kind of data under analysis. In case of dichotomous dependent variables that are nominal or ordinal by nature, logistic regression models are used. Since the baseline risks or survivor functions are not stated, PH and AFM are truly semiparametric models of regression analysis. But if the maintaining probability distributions are given, the accelerated models may also be parametric in nature [13].

Regression models for survival data are employed when conducting statistical analysis of data that has an observational measure made up of the time until an event of interest occurs. In survival analysis, the method can be used to several supplementary situations, but typical ones are failure, relapse, or death. For survival data, the Cox Proportional, Hazards model is accompanied by parametric survival models as the main types of regression models. Proportional hazard rate (PH) model is defined as follows:

$$\hbar(t|X_i) = \hbar_0(t).exp(\beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p)$$

Where $\hbar_0(t)$ define as baseline hazard, $\beta_1, \beta_2, \ldots \beta_p$ are coefficient regression of independent variable $X_1, X_2, \ldots, X_p$.

A parametric model is similar to the semi-parametric model with only one difference: the distribution of the survival time is known in the parametric model, but in the semi-parametric analysis, emphasis is laid on covariates rather than risk factors. Furthermore, the specification of the parametric model is in contrast with the other models including the non-parametric, and the semi-parametric since the former is capable of determining the distribution of existence by utilizing the full maximum likelihood in estimating the parameters, the use of residuals to measures the variation between the observed and the estimated survivor's time, as well as, generate clinically useful estimations from the recognized parameters [14, 15]. The failure time model with an accelerated rate is:

$$Ln(T_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \eta\varepsilon_i$$
$$i = 1,2,\ldots,n$$

Where η is scale parameter, ε is error term follow a specific distribution. T_i represent the failure time logarithm, it is composed of $\beta_1 X_1, \beta_2 X_2, \ldots, \beta_p X_p$[13, 16, 17].

3. Weibull AFT

The Weibull accelerated survival model is known to include an exponential base distribution, which is the age period hazard rate (APH) model. The Weibull accelerated survival model is used to describe the conditional distribution of age at death, given the age at which the individual becomes deceased. Furthermore, the Weibull accelerated survival model assumes a scaled Weibull density as the baseline hazard function, with hazard rate $\lambda_0(t)$ when the age at death equals failure time τ conditional on surviving beyond age t [18]. In particular Weibull AFT model is useful in performing relationship analysis between the variables of interest and the time until the event occurrence of interest utilizing

the Weibull distribution [17]. Due to its versatility in accommodating different forms of failure time data, and capability to provide various hazard functions, exponential decay Weibull distribution is one of the survival analysis functions.

Sets the shape parameter and the scale parameter continue to demonstrate the relationship at this Weibull AFT. The one which is in front of it is the shape parameter which is assuming the constant attribute. The AFT model which is defined as a transformation of the response variable from logarithmic or monotone, the time of failure. This model is a type of real linear regression model [20 ،19] .

The probability density function (P.D.F.) of Weibull with two parameters is defined as follows [21, 22]:

$$f(t;\alpha,\lambda)=\frac{\left(\frac{\lambda}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\lambda}e^{-\left(\frac{t}{\alpha}\right)^{\lambda}}}{\left(\frac{t}{\alpha}\right)}$$

where $\alpha>0\,,\lambda>0\,,t\geq 0$

Where λ and α are the scale and shape parameter respectively. the cumulative distribution function (CDF) will be [23]:

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

For the Weibull distribution, the survival function is provided by [24, 25]:

$$\delta(t)=1-F(t)$$
$$=1-\left(1-e^{\left(\frac{-t}{\alpha}\right)^{\lambda}}\right)=e^{\left(\frac{-t}{\alpha}\right)^{\lambda}}$$

The hazard function depends on the value of α [18]:

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

But for different α we get different hazards for α as follows [26]:

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

The instantaneous failure rate, or hazard function, increases with time when $\alpha>1$. This kind of things are typical for so called bathtub curve where failure rates are low initially and then they gradually increase and stabilize again. When, $\alpha<1$, then the hazard function reduces over time, thus implying a declining failure rate. The distribution becomes the exponential distribution when α=1 the hazard function remains constant over time [27]. The scale parameter λ whose function is to influence the spread or scale of such a distribution is another factor. The greater values of λ mean more shrinkage in the horizontal direction, so obtaining shorter times. After analyzing the above hypothetical data sets, we observe that the values of λ closer to 1 stretch the distribution and hence the longer durations.

Now suppose that:

$$ln(T)=X\beta+\sigma\varepsilon$$

Here ε follow standard Gumbel distribution with mean=0 and variance=1. let $\lambda=e^{-X\beta}$, then Eq. (3), (4), (5) and (6), will be respectively [17]:

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

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

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

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

The Weibull AFT model's right part of the equation is the linear predictor that ties the scale parameter $\lambda=e^{-X\beta}$ to the variables.

The coefficients relate the variables to the logarithm of the survival time and, therefore, to the scale parameter of the Weibull distribution $\beta_0,\beta_1,\ldots,\beta_p$. It may thus determine how each covariate affects the scale parameter and, therefore, scale survival time in exponentiating the coefficients. Weibull AFT models are also used often in survival analysis when the distribution of survival times is expected to follow a Weibull distribution, and factors that contribute to the time of an event are of interest. These models are useful because

they demonstrate how the various components connect as well as the rate at which an occasion progresses [28].

4. Lemur's Optimization Algorithm

Gorrok Lemurs Optimization Algorithm (Lemurs) is a population-based evolutionary strategy that was introduced by Bianchi et al. in 2017. The main advantage of the Lemurs algorithm is that it can be easily adapted to solve complex and structured optimization problems, due to the functions that allow the exchanges and transfers of genetic information among the metaheuristic components incorporated in the Lemurs. In addition to its scalability and flexibility, Lemurs also implements a stochastic batch semantics-based selection method. This section conservatively reviews the key components of the original Lemurs optimization algorithm. More background of these components will be presented in the following sections regarding the component's composition [29].

Lemurs are placed among the prosimians and all members of this group are primates (Kappeler and van Schaik 2002). These are rodents of many types and are found in some countries like Madagascar and some parts of the country of Comoros. They are inhabitants of mountains, swampy, forest, rain, thorn and deciduous regions and forests. Lemurs are of kinds; the largest kind is the Andri. It weight can go up to 15kg. Customers range in weight from 7 to 10kg and length from 60 to 90cm. The least one is the Murid of Madame Berthe that has a weight of 3g and size of 9-11 cm in length [30].

Lemurs are social animals that are predominantly gregarious and group animals that mostly assemble in troops where common sexuality is dominant. Female is also larger than the ring-tail lemur and their population totals from 6 to 30 species. Lemurs sleep a considerable part of each day then spend most of their waking time in trees. Based on the research, since communication in animal species is basically passing of information, there are two distinct ways the Lemurs utilize in their communication. They also employ sound by vocalizing, and by emitting chemical substances in form of odors.

For the LO algorithm, we used two main lemur behaviors as inspiration: jump and dance-hup. The lemurs also leap upwards and sit on a branch and grabbing the trunk with both the limbs of the hand and the feet. It is possible for them to jump as far as between 10 meters from one tree trunk to another in few moments. The dance-hup is performed when distances between trees are large; the lemurs travel over distances greater than one hundred meters two-legs and by leaping sideway with their arms extended, and their limbs sway from chest to head high, it is believed to aid balance [31].

5. Models and Methods

The search process is divided into two phases in the population-based algorithm, as described in the previous section: and so, it was that some published essays on exploration versus exploitation. The dance-hup behaviour is used in the exploration phase of this framework. In contrast, the leap-up behaviour is favourable to LO in the sense that, through it, the search space is broadened. This view means that every solution is a lemur while every vector must be one of the coordinates of the lemur in question. We also propose the best position to each solution that has a relationship with the fitness function value of the solution. Thus the lemurs either update their place vectors and dance hup to local best neatest or leap up to the location of no other lemur is better than the global best lemur. This part provides provisions of LO algorithm formulas in Mathematics, the flowchart and steps of the algorithm. LO is considered one of the powerful population-based algorithms so, the lemurs set is represented in matrix form In the following equation itself the matrix of input population for the LO algorithm is defined.

$$x = \begin{bmatrix} l_1^1 & l_1^2 & \dots & l_1^d \\ l_2^1 & l_2^2 & \dots & l_2^d \\ \vdots & \vdots & \vdots & \vdots \\ l_n^1 & l_n^2 & \dots & l_n^d \end{bmatrix}$$

where n denotes the matrix of the algorithm set, of size n×d; n represents the candidate solution and d the decision variable. To use the LO for solving an optimization problem like Feature selection (FS), the function of the LO algorithm runs into many steps:

Step (1): Describe the following Lemurs parameters: N Population – the number of individuals in the data set. Max_{iter} means the number of iterations. d reflects the natural proportionality between the dimensionality of the space being searched and the sizes of the data sets. Moreover, UB is the upper-bound while LB is the lower bound.

Step 2: Generate x decision variable in ith solution based on Eq (2):

$$x_i^j = \left(LB + \left(UB_j - LB_j\right)\right) \times r$$

where r refers to the uniform random number $\in [0,1]$.

Step 3: Compute Free Risk Rate (FRR):

$$FRR = HRR - t \times \left((HRR - LRR)/Max_{iter}\right)$$

Where:

t: iteration number.

Max_{iter}: iteration size.

HRR: high risk rate

LRR: low risk rate

Step 4: compute the fitness value for each x_i^j, as expressed in the following equation:

$$Fit\left(x_i^j\right) = \alpha \times (1 - Acc) + \beta \times (s/S)$$

where $Fit\left(x_i^j\right)$ is the fitness value, s denoted to the total of selected features, S is the maximum selected feature and Acc is the accuracy of each subset (when each subset is evaluated in every iteration, it is the accuracy score of each subset which is determined by the KNN classification function).

Step 5: In order to increase the value of the fitness of the lemurs, we sort into two different procedures. Firstly, we apply the best near lemurs (bnl) approach which in other words corresponds to the identification of the solutions with minimal fitness value. According to the adopted FS objectives, bnl will be endowed with the best features for the current development iteration. Second, we choose the global best lemur, abbreviated as gbl, from all the lemurs within the population, which depicts the global best solution.

Step 6: Define r_1 which is a random number $\in [0,1]$ and comparing it with FRR. Then, update the position for each lemur

$$x_i^j = \begin{cases} x(i,j) + \left|x(i,j) - x(bnl,j)\right| \times (r_1 - 0.5) \times 2 \\ r_1 < FRR \\ x(i,j) + \left|x(i,j) - x(gbl,j)\right| \times (r_1 - 0.5) \times 2 \\ r_1 > FRR \end{cases}$$

The current i^{th}lemur of N^{th}population is $x(i,j)$, being the candidate solution in the j^{th} decision variable wherein bnl represent the best near lemurs, the present solution in this iteration of the entire algorithm and (gbl) is the global best lemurs for the whole population across all iterations is described. Then it attempts to approach the two or more lemurs with less favourable fitness values in the round through the dance hup, as a more favourable fitness value. The optimization procedure is begin randomly to form the set of lemurs which influences. The FRR value is set just below the LRR, which means that the lemur begins moving and moves towards acquiring the best nearest one using the action 'dance hup'. Now, the function of LO performs this dance hup action in order to reduce the FRR to HRR value. Then it uses the leap up action to move the lemur closer to the global optimum solution Then, it uses the leap up action. This action is done to the last stopping condition is met, in other words until the last signal is detected. Some of the movement behaviours of lemurs have been illustrated as follows; Leap up and Dance hup shown in Fig.1.

6. The Proposed approach

Nonlinear chaotic maps in chaotic systems are important in engineering, biology, and economics due to such properties as ergodicity, mixing property, and sensitivity to initial conditions. In this paper, to perform the feature selection, we utilize Support Vector machines based on Statistical analysis technique known as LO.

An optimisation problem with a search on the interval [0,1]: With regards to improving the LO performance, chaotic maps are considered in the paper. It is postulated that

chaotic maps perform LO to liberate solutions from local traps in getting a steep convergence of the selection of variables in the Weibull model.

Next, ten chaotic maps are used to manipulate the random parameter values of LO in this paper.

Since the initial values can affect the fluctuation pattern, we have normalized the initial point for all the chaotic maps to point 0.7. As for the other parameters we have left them unchanged.

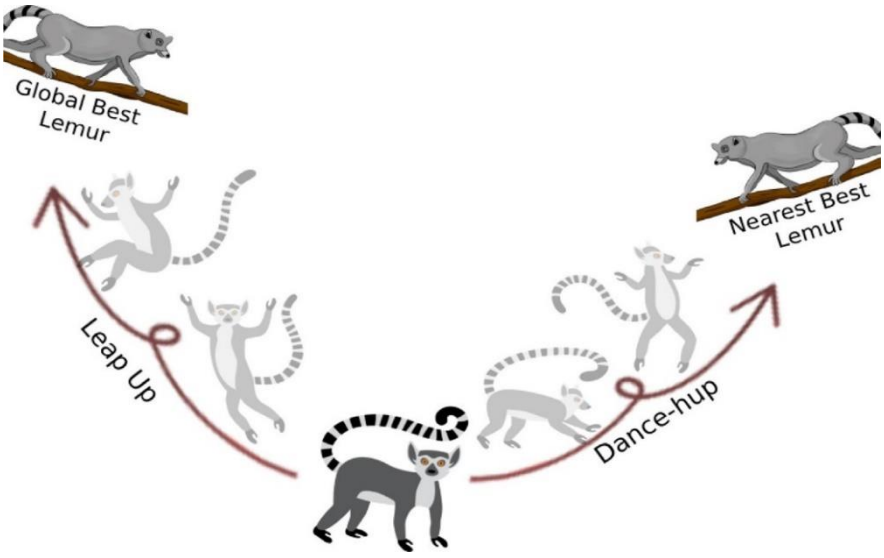

Figure 1. Lemurs' behaviour movement.

Any of the features could be declared as binary decision variables that denote the extent of the particular feature in the concept of the model [32]. It is well understood that for the sake of the present consideration let the vector that has D elements denote the entire feature set, The vector model is where any element of the vector represents the feature, and if this feature is selected, the number '1' is assigned; if the feature is not selected, '0' is assigned. Hence, the LO approach will fit into the continuous space which defines the feature selection as an optimization problem rather than the discrete space. Knowing this we have to assume that S is discrete. Therefore, the configuration of our suggested method is as follows:

Step (1): Determine the following: Number of Lemur $\eta_c = 40$, the randomization parameter $\alpha = 0.1$ and the upper limit of iteration t_{max}.

Step (2): The Lemur positions in the original LO are produced follow a continuous uniform distribution on interval [0,1]. The proposed chaotic maps utilise the maps outlined in Table 1.

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

$$fitness = min\left[\frac{1}{n}(\sum_{i=1}^{n}(y_i - \hat{y}_i)^2)\right] \ldots(17)$$

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

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

In order to compare how accurate the predictions of the models using the AFT approach are, we adopted several of regularization techniques through simulation studies which include LO, Elastic net, Lasso, L1/2 and MCP [33]. The AFT model simulation schemes used were modelled after Bender's work. The following is how our simulation data were produced [34]:

Table 1: 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	$x_{K+1} = mod\left(x_K + 0.2 - \frac{0.5}{2\pi} sin(2\pi x_K), 1\right)$	(0,1)
Guass/mouse	$x_{K+1} = \begin{cases} 1 & x_K = 0 \\ \frac{1}{mod(x_K, 1)} & otherwise \end{cases}$	(0,1)
Iterative	$x_{K+1} = sin\left(\frac{(0.7)\pi}{x_K}\right)$	(-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	$x_{K+1} = 1.07(7.86x_K - 23.31(x_K)^2) + 28.8(x_K)^3 - 13.30288(x_K)^4$	(0,1)
Sinusoidal	$x_{K+1} = 2.3x_K\, sin(\pi x_K)$	(0,1)

Tent $$x_{K+1} = \begin{cases} \frac{x_K}{0.7} & x_K < 0.7 \\ \frac{10}{3}(1 - x_K) & x_K \geq 0.7 \end{cases} \quad (0,1)$$

1. Determine the correlation coefficient ρ and build an array $\delta_0, \delta_{i1}, \delta_{i2}, \ldots, \delta_{ip}$, $i = 1,2,\ldots,n$ are independent following standard normal distribution, and set:

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

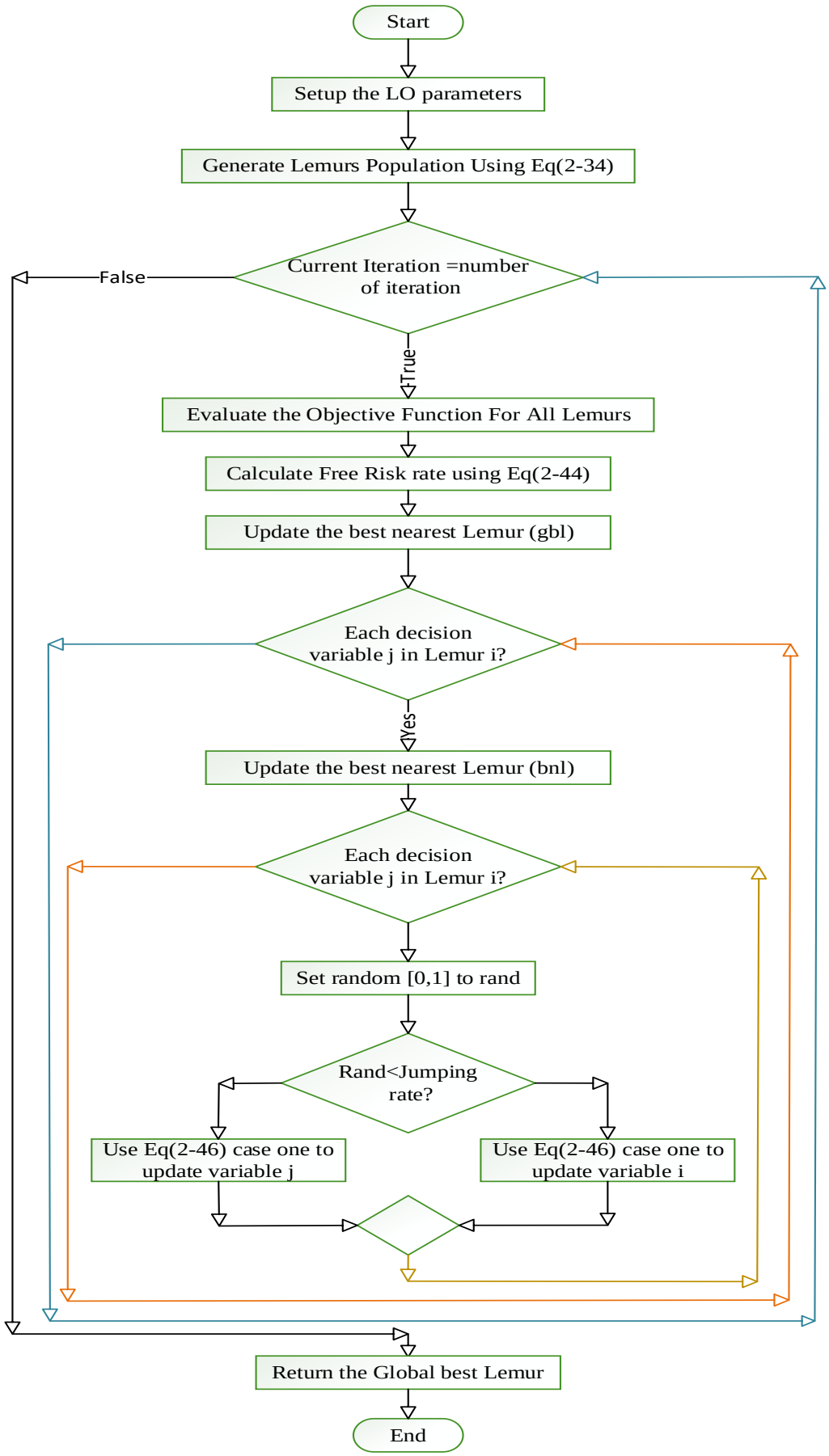

Figure 2: Lemur Flow Chart

2. The dependent Survival time:

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

3. Suppose that y_i' follow a random distribution where $i = 1,2,\ldots,m$ and m is the number of the censored sample. The number of censored data is determined by the censoring rate.
4. Let $y_i = min(y_i, y_i')$, 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.
5. Since our simulation rely on Leukaemia cancer data, we define 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 $\beta_{13} = 1.3$ while the other 992 genes are equal to zero. It is also good to note that x_1, x_2, x_3, x_5, x_9, x_{10}, x_{11} and x_{13} are significant variables in this analysis. The right censored (C) value is fixed at 10%, 20% and 40%, training sample size is dependent on three values as $n = (100,300,500)$ the correlation coefficient is $\rho = 0.4$. As technique of to optimize fidelity, a model for regularization regulation can be examined on to the 50 data sets for a training set and to get a prediction on the 50 data set for the testing set. For example, each of those outcomes was obtained by dividing the 200 out-turn as the fruit of opening 200 doors. They are, the total number of features and correct number of features skewed by each of the explored regularization methods. These 200 are the results reached at the end of two hundred consecutive experiments done successively. To be clearer, primary outcomes mean average of total features and total selected features as well as the number of correct features are given in the Table (2) for the candidates of a beach regularization approach accompanied by 200 repeat tests. However, when the training sample size is extremely limited (n=50), all of the techniques were seen to be very laborious in the precision of the features that define the right genes. The probability of recognizing accurate non-zero features increases as ‘n’ value increases. A reflective account of the outcomes accomplished by the LO maps is conducted. Moreover, the performance measure of the proposed system is compared with the current LO with all its parameters. In this case an assessment of the quality of the performance done by using mean-squared error (MSE) in both the training and testing databases coupled with the number of chosen variables. The detail on the identified themes is presented in the following tables 3, 4 and 5 below:

Table 2: A simple example

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

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
	$C = 10\%$		$C = 20\%$		$C = 40\%$	
LO	26	4.200	26	5.501	26	7.106
Chebyshev	22	3.923	22	5.153	22	6.853
Circle	19	3.666	19	4.766	19	6.666
Guass	21	4.093	21	5.363	21	7.143
Iterative	24	3.923	24	5.173	24	6.873
Logistic	25	4.253	25	5.553	25	7.354
Piecewise	18	3.406	18	4.536	18	6.236
Tent	16	3.188	16	4.488	16	6.188
Singer	12	2.984	12	3.684	12	5.384
Sinusoidal	13	3.122	13	4.167	13	5.867
Sine	10	2.666	10	3.866	10	5.566

Table 4: Implementing the methods used to train data when $n = 300$

Method	No. of selected variable	MSE	No. of selected variable	MSE	No. of selected variable	MSE
	$C = 10\%$		$C = 20\%$		$C = 40\%$	
LO	26	3.210	26	4.183	26	2.746
Chebyshev	22	2.973	22	3.873	22	2.509
Circle	19	2.666	19	3.766	19	2.202
Guass	21	2.793	21	3.693	21	2.329

Iterative	24	2.883	24	3.783	24	2.419
Logistic	25	3.243	25	4.265	25	2.779
Piecewise	18	2.476	18	3.346	18	2.012
Tent	16	2.088	16	2.988	16	1.624
Singer	12	1.994	12	2.894	12	1.53
Sinusoidal	13	2.14	13	2.951	13	1.676
Sine	10	1.776	10	2.676	10	1.312

Table 5: Implementing the methods used to train data when $n = 500$

Method	No. of selected variable	MSE	No. of selected variable	MSE	No. of selected variable	MSE
	$C = 10\%$		$C = 20\%$		$C = 40\%$	
LO	26	2.211	26	2.911	26	3.824
Chebyshev	22	1.974	22	2.874	22	3.09
Circle	19	1.767	19	2.467	19	2.883
Guass	21	1.894	21	2.994	21	3.01
Iterative	24	1.984	24	2.684	24	3.1
Logistic	25	2.264	25	2.964	25	3.48
Piecewise	18	1.477	18	2.177	18	3.883
Tent	16	1.089	16	1.789	16	2.205
Singer	12	0.995	12	1.795	12	2.111
Sinusoidal	13	1.191	13	1.891	13	2.267
Sine	10	0.777	10	1.477	10	1.893

Furthermore, the Sine map achieves the lowest Mean Squared Error (MSE) compared to the other chaotic maps utilized. Compared to the LO, the Sine map exhibited a decrease of approximately 36.667%- 21.67%, 44.672%-52.221% and 64.857%-50.496% from tables 3, 4 and 5 respectively.

7. Real Application

For performance analysis of the proposed strategy, the recommendation is made to run the experiment on the gene's dataset of a real-world case. The content, datasets, and justification for this study are summarized in table 5. The first of them is the selection of the Disseminated Large B-cell Lymphoma dataset

(DLBCL) [35] as the source of material. In total, the study involves 240 samples from patients with lymphoma. For each patient, we have 7399 gene expression values and the corresponding survival time, which could be censored. The second dataset is the Dutch breast cancer dataset, also known by the abbreviation DBC, which has information on 79 patients and thirty independent variables. This is a data set of the 295 treatments given to the breast cancer patients. The raw measurement for every patient consists of 4919 gene expression values. They also obtained RNA-Seq signatures of the tumors of the patient [36]. The third class is cancer in the lung, abbreviated as LC. The data comprises of 86 patients diagnosed with lung cancer. The expression level data of that particular gene is 7129 and it also contain survival time for example alive or censored [37].

Akaaki's criterion (AIC) and Bayes' criterion (BIC) [38] will be used as methods for selecting variables according to formulas [39]:

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

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

Where k equal to the number of explanatory variables.

In the following, we will demonstrate the efficiency of the proposed model comparing it to AIC and BIC. The step to execute this is to randomly split the expressions of genes in each dataset into a training and an unseen dataset with 70% data allocated to the training dataset and the rest 30% allocated to the testing dataset. Here we use the time-dependent receiver-operator characteristics curves for the censored data in evaluating our algorithm for the prediction of the results in Table 6 highlighting the fact that world application is the main result in real-life results.

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

Table 6 shows the average of the three realistic datasets used and the approaches employed to realize them. This evidence is enough to infer that, given that AIC selects many more genes than the BIC and the LO, there already exist huge differences between the three approaches. Out of the 3 applications, the LO then selected the minimum number of genes in the ultimo subset. This is useful for the model performance evaluation where the mean AUC of both types of data is given to make clear distinction and to show the

correlation between both the training and testing sets. Tables 3 and 4 below show how they meet the objective which I defined as a standpoint for interval measurements and a perceived distance. The shown results proved that Sine map with accuracy levels of 97.2% for DLBCL, 97.6% for DRBC, and 98.1% for LC datasets was

recognized as the highest precision as compared to map Singer with the percentage of 95.2% for DLBCL, 97.9% for DRBC and 98.2% for LC, Finally, the tool to quantify the performance, Therefore, it could be safely concluded that the supremacy of the multilayer perceptron is due to the fact that the AIC and BIC computed on, almost in the same manner as across, the datasets did not make different performances. If so, it should be evidenced that LO's "contribution" will be better than AIC's.

Table 7: The AUC results for the training dataset

	DLBCL	DBC	LC
AIC	0.863	0.868	0.870
BIC	0.871	0.884	0.951
LO	0.910	0.912	0.930
Logistic	0.917	0.917	0.941
Iterative	0.901	0.911	0.942
Chebyshev	0.907	0.928	0.947
Guass	0.916	0.936	0.956
Circle	0.918	0.941	0.957
Piecewise	0.934	0.940	0.961
Tent	0.926	0.960	0.969
Sinusoidal	0.933	0.958	0.966
Singer	0.948	0.971	0.978
Sine	0.972	0.976	0.981

Table 8: The AUC results for the testing dataset

	DLBCL	DBC	LC
AIC	0.725	0.774	0.780
BIC	0.748	0.799	0.861
LO	0.771	0.818	0.849
Logistic	0.773	0.821	0.857
Iterative	0.779	0.827	0.858
Chebyshev	0.782	0.848	0.867
Guass	0.798	0.845	0.866
Circle	0.800	0.847	0.870
Piecewise	0.808	0.856	0.872
Tent	0.817	0.867	0.876
Sinusoidal	0.869	0.871	0.897
Singer	0.888	0.874	0.931
Sine	0.921	0.951	0.978

This was an indication that other maps; Sine, Singer, Sinusoidal, Tent, Piecewise, Circle, Gauss, Chebyshev, Iterative and Logistic maps were respectively accurate in identifying close ends with probability greater than 0.95 who actually had a chance of having cancer.

7. Conclusion

Censoring influences the selection variables of the quality biases in the Weibull AFT compliance models. They applied a Lemur optimization of algorithms with ten architectural styles, which could be an algorithm for feature selection. The results of simulations and real data illustrate the effectiveness of identical respective algorithms out of a swarm of LO in MSE training data. Furthermore, it showed an importance in this case that is greater than all the other factors in this regard.

References

1. Clark, T.G., et al., *Survival analysis part I: basic concepts and first analyses.* British journal of cancer, 2003. **89**(2): p. 232-238.
2. Khanal, S.P., 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, 2014. **3**(6): p. 161-6.
3. Hassan, M.K. and H.T. Abed, *Comparison of survival models to study determinants liver cancer.* journal of Economics And Administrative Sciences, 2021. **27**(125).
4. Lambert, P., et al., *Parametric accelerated failure time models with random effects and an application to kidney transplant survival.* Statistics in medicine, 2004. **23**(20): p. 3177-3192.
5. Kim, B.S. and G. Maddala. *Estimation and specification analysis of models of dividend behavior based on censored panel data*. in *Panel data analysis*. 1992. Springer.
6. Kalbfleisch, J.D. and R.L. Prentice, *Estimation of the average hazard ratio.* Biometrika, 1981. **68**(1): p. 105-112.
7. Gelfand, L.A., et al., *Mediation analysis with survival outcomes: accelerated failure time vs. proportional hazards models.* Frontiers in psychology, 2016. **7**: p. 423.
8. Abolghasemi, J., et al., *Survival analysis of liver cirrhosis patients after transplantation using accelerated failure time models.* Biomedical Research and Therapy, 2018. **5**(11): p. 2789-2796.
9. Moran, J.L., et al., *Modelling survival in acute severe illness: Cox versus accelerated failure time models.* Journal of Evaluation in Clinical Practice, 2008. **14**(1): p. 83-93.
10. Andersen, C.R., et al., *Accelerated failure time survival model to analyze Morris water maze latency data.* Journal of neurotrauma, 2021. **38**(4): p. 435-445.
11. Weibull, W., *A statistical distribution function of wide applicability.* Journal of applied mechanics, 1951.
12. Karimi, A., 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, 2016. **12**(3): p. 1184-1188.
13. Chai, H., 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, 2015. **64**: p. 283-290.
14. Hosmer, D.W. and S. Lemeshow, *Survival analysis: applications to ophthalmic research.* American journal

of ophthalmology, 2009. **147**(6): p. 957-958.
15. Wang, H., H. Dai, and B. Fu, *Accelerated failure time models for censored survival data under referral bias.* Biostatistics, 2013. **14**(2): p. 313-326.
16. Huang, J., S. Ma, and H. Xie, *Regularized estimation in the accelerated failure time model with high-dimensional covariates.* Biometrics, 2006. **62**(3): p. 813-820.
17. Liu, E., R.Y. Liu, and K. Lim, *Using the Weibull accelerated failure time regression model to predict time to health events.* Applied Sciences, 2023. **13**(24): p. 13041.
18. Rasheed, H., 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, 2012. **53**(2): p. 362-366.
19. Kareem, S.A., *A Comparison Between Two Shape Parameters Estimators for (Burr-XII) Distribution.* Baghdad Science Journal, 2020. **17**(3 (Suppl.)): p. 0973-0973.
20. Gorgoso-Varela, J.J. and A. Rojo-Alboreca, *Use of Gumbel and Weibull functions to model extreme values of diameter distributions in forest stands.* Annals of forest science, 2014. **71**: p. 741-750.
21. Abdulwahab, R.A., 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, 2014: p. 1120-1126.
22. Alkanani, I. and D. Majeed, *Parameters Estimation for Modified Weibull Distribution Based on Type One Censored Samplest.* Iraqi Journal of Science, 2013. **54**(Mathematics conf): p. 822-827.
23. ALNaqeeb, A. and A. Hamad, *Suggested method of Estimation for the Two Parameters of Weibull Distribution Using Simulation Technique.* Iraqi Journal of Science, 2013. **54**(Mathematics conf): p. 744-748.
24. Juhan, N., et al., *Comparison of stratified Weibull model and Weibull accelerated failure time (aft) model in the analysis of cervical cancer survival.* Jurnal Teknologi, 2016. **78**(6-4).
25. Karim, M.R. and M.A. Islam, *Reliability and survival analysis*. 2019: Springer.
26. Jaafar, S.M., B.I. Abdulwahhab, and I.A. Hasson, *Best estimation for the Reliability of 2-parameter Weibull Distribution.* Baghdad Science Journal, 2009. **6**(4): p. 705-710.
27. David, G.K. and K. Mitchel, *Survival analysis: a Self-Learning text*. 2012, Spinger.
28. Elangovan, R., R. Mohanasundari, and E. Susiganeshkumar, *ACCELERATED FAILURE TIME MODEL FOR SURVIVAL ANALYSIS.* Journal of Information and Computational Science, 2020. **9**: p. 10-33.
29. Abasi, A.K., et al., *Lemurs optimizer: A new metaheuristic algorithm for global optimization.* Applied Sciences, 2022. **12**(19): p. 10057.
30. Radespiel, U., et al., *Sex-specific usage patterns of sleeping sites in grey mouse lemurs (Microcebus murinus) in northwestern Madagascar.* American Journal of Primatology, 1998. **46**(1): p. 77-84.
31. Powzyk, J. and C. Mowry, *The Feeding Ecology and Related Adaptations of Indri indri. Lemurs, 353–368*. 2006.
32. Mahmood, R.A.R., A. Abdi, and M. Hussin, *Performance evaluation of intrusion detection system using selected features and machine learning classifiers.* Baghdad Science Journal, 2021. **18**(2 (Suppl.)): p. 0884-0884.
33. Alkhateeb, A.N. and Z.Y. Algamal, *Variable selection in gamma regression model using chaotic firefly algorithm with application in chemometrics.*

Electronic Journal of Applied Statistical Analysis, 2021. **14**(1): p. 266-276.
34. Sayed, G.I., A. Darwish, and A.E. Hassanien, *A new chaotic whale optimization algorithm for features selection.* Journal of classification, 2018. **35**: p. 300-344.
35. Rosenwald, A., et al., *The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma.* New England Journal of Medicine, 2002. **346**(25): p. 1937-1947.
36. Van Houwelingen, H.C., et al., *Cross-validated Cox regression on microarray gene expression data.* Statistics in medicine, 2006. **25**(18): p. 3201-3216.
37. Beer, D.G., et al., *Gene-expression profiles predict survival of patients with lung adenocarcinoma.* Nature medicine, 2002. **8**(8): p. 816-824.
38. Awad, M.K. 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, 2020: p. 1433-1439.
39. Vrieze, S.I., *Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC).* Psychological methods, 2012. **17**(2): p. 228.