ISSN E.2958-6941 ,ISSN P.1818-1074 Vol. 22 issue88. June. 2026

http://journals.uokerbala.edu.iq/index.php/ijas/index
Iraqi Journal for Administrative Sciences
المجلة العراقية للعلوم الإدارية

Performance Evaluation of Machine Learning & Robust Estimation for Parameters in Survival Analysis[1]

Hozan Taha Abdalla[1]* Samira Muhamad Salh[2]

تقييم أداء تعلم الآلة والتقدير الحصين للمعلمات في تحليل البقاء

هوزان طه عبد الله*[1] ، سميرة محمد صالح[2]

1. *Department of Statistics and Informatics, College of Administration and Economics, Sulaimani University, Sulaimani, Kurdistan Region – Iraq. hozan.abdulla@univsul.edu.iq
2. Department of Statistics and Informatics, College of Administration and Economics, Sulaimani University, Sulaimani, Kurdistan Region – Iraq samira.muhamad@univsul.edu.iq

1. قسم الإحصاء والمعلوماتية، كلية الإدارة والاقتصاد، جامعة السليمانية، السليمانية، إقليم كردستان – العراق
2. قسم الإحصاء والمعلوماتية، كلية الإدارة والاقتصاد، جامعة السليمانية، السليمانية، إقليم كردستان – العراق

Article information

Abstract DOI: https://doi.org/10.71207/ijas.v22i88.6017

The goal of survival analysis as a branch of statistics is to analyze data, model it and estimate the time it takes for a certain event to occur. The event outcomes for the existence instances are not observable after a specific amount of time or instances that do not experience any events in the period of observation present one of the main challenges in this case. The best way to deal with this-so called censoring is to use survival analysis techniques. Furthermore, numerous algorithms of machine learning have been modified to handle this type of censored data as well as other difficult issues that come up in real world data. In this paper we provide a comprehensive comparison between parametric survival models by using Buckley James estimation method, Robust parametric survival models by using Buckley James estimation method with Tukey's Biweight function and machine learning method to build a survival tree for (Exponential, Weibull, Weibull Three Parameter and Log-normal) parametric models. Based on these comparisons the best models were detected by depending on the smallest AIC and BIC.

Articlehistory: DD/MM/YY
Received: 04/04/2026
Accepted: 22/04/2026
Available online: 03/06/2026

Keywords: Machine Learning, Survival Analysis, Censoring, Robust, Parametric Models

تاريخ الاستلام:2026/04/04
تاريخ قبول النشر: 2026/04/22
تاريخ النشر: 2026/06/03

الكلمات المفتاحية: التعلم الآلي، تحليل البقاء، الرقابة، النماذج الحصينة، النماذج البارا مترية.

Citation: Taha Abdalla, Hozan, Muhamad Salh, Samira. (2026). Performance Evaluation of Machine Learning & Robust Estimation for Parameters in Survival Analysis, *Iraqi Journal for Administrative Sciences*, 22(88), 87-104.

الاقتباس: طه عبد الله، هوزان، محمد صالح، سميرة. (2026). تقييم أداء تعلم الآلة والتقدير الحصين للمعلمات في تحليل البقاء، *المجلة العراقية للعلوم الإدارية*، 22(88)، 87-104.

المستخلص

إن الهدف من تحليل البقاء -والذي هو فرع من فروع الإحصاء- حيث يقوم بتحليل البيانات والنمذجة و تقدير الوقت الذي يستغرقه حدوث حدث معين؛ أذ تعتبر نتائج الأحداث للحالات الموجودة غير قابلة للمراقبة بعد فترة زمنية محددة أو عند الحالات الخالية من أية أحداث خلال فترة المراقبة من التحديات الرئيسة في هذا السياق؛ حيث أن أفضل طريقة للتعامل مع هذا تسمى المراقبة وهي : استخدام تقنيات تحليل البقاء. علاوة على ذلك، تم تعديل العديد من خوارزميات تعلم الآلة للتعامل مع هذا النوع من البيانات المقطوعة وكذلك المشكلات الصعبة الأخرى التي تظهر في البيانات الحقيقية. نقدم في هذه الورقة مقارنة شاملة بين نماذج البقاء البارامترية باستخدام طريقة تقدير (Buckley James) ، ونماذج البقاء البارامترية القوية باستخدام طريقة تقدير (Buckley James) مع دالة (Tukey's Biweight)، وطريقة التعلم الآلي لبناء شجرة البقاء للنماذج البارامترية (Exponential, Weibull, Weibull Three Parameter and Log-normal). استنادًا إلى هذه المقارنات، تم اكتشاف أفضل النماذج بالاعتماد على أصغر قيم AIC و BIC.

---

[1] Extracted from a doctoral dissertation: Comparison among Robust Estimation Methods of Survival Functions for Exponential Family Distribution.

1. Introduction

The main goal of the majority of these real world implementation is to improve the estimation of time at which a specific event of interest take place. One of the primary issues with such time to event data is the prevalence of censored cases, in which the event of interest is not distinguished because of time constraints or the observer lost track during the scrutiny period. Frequently, an event occurs in one instance, but data regarding the outcome variable of the other cases is only obtainable until some point in time. Because of this, directly apply predictive algorithms based on the statistical and machine learning methods typically utilized to analysis survival data is not feasible. An essential area of statistics is survival analysis, which offers a number of solutions for censored data issues that come up during complex data modeling. When modeling a specific event of interest is the primary goal, this is also known as time to event data, and it is widely used in real world implementation domains. Although typically the statistical techniques are designed to deal with data which is low-dimension, machine learning techniques are typically applied to high-dimensional problems. Because by using both statistical and machine learning methods survival problems can be analyze, using machine learning techniques provide more effective algorithms for survival analysis. This allows them to take advantage of recent advancements in optimization and machine learning to know how covariates are dependent on the survival times in various ways. The classical statistical procedures can be classified into three groups by depending on the assumptions made and the model's use of parameters: (a) non-parametric models, (b) semi-parametric models, and (c) parametric models. A distinct branch includes machine learning algorithms such as support vector machines, neural networks, Bayesian techniques, and survival trees, all of which have grown in popularity recently (Wang et al., 2019).

In this paper Buckley James Estimator is used with parametric Accelerated Failure Time (AFT) model to censored survival data with (Exponential, Weibull, Weibull Three Parameter and Log-normal) distributions, using an iterative imputation approach combined with bootstrap, and Robust Buckley James Estimator with Tukey's Biweight function is used with parametric Accelerated Failure Time (AFT) model to censored survival data with (Exponential, Weibull, Weibull Three Parameter and Log-normal) distributions, using an iterative imputation approach combined with bootstrap and using machine learning method to build a survival tree by using bootstrap for parametric models (Exponential, Weibull, Weibull Three Parameter and Log-normal).

2. Literature Review

Interest in the development of machine learning and survival analysis methods has grown in recent years. The following review investigates previous studies that have focused on machine learning and survival analysis.

(Boček & Štěpánek, 2025) This study by Boček and Štěpánek (2025) offers a thorough comparison of six survival analysis models, ranging from traditional to contemporary machine learning techniques, applied to both simulated datasets with known ground truth and real-world ICU patient data from the MIMIC-III database. Random Survival Forests (RSF), Conditional Inference Forests (CIF), Oblique Random Survival Forests (ORSF), CoxBoost, penalized Cox regression with elastic net regularization, and the standard Cox PH model are among the models evaluated. Harrell's concordance index was used to assess performance for different sample sizes and censoring circumstances. The findings show that tree-based ensembles were consistently outperformed by boosting and penalization-based techniques (CoxBoost and elastic net-regularized Cox), especially when there was high-dimensional noise or little data.

(Kaindal & Venkataramana, 2025) Six parametric distributions (Weibull, exponential, Gaussian, logistic, log-logistic, and log-Gaussian) and five machine learning algorithms (neural networks, SVM, random forest, GBM, and logistic regression) are evaluated by Kaindal and Venkataramana (2025) using SEER data from 2,085 ILC patients. The best balance between complexity and fit was shown by random forests, as indicated by the smallest AIC/BIC values, while neural networks obtained the highest predictive accuracy. Radiation therapy, age, tumor grade, AJCC stage, and marital status were consistently identified as important prognostic factors. This comparative analysis promotes multi-metric evaluation frameworks to maximize survival prediction in oncology by highlighting the complementary strengths of parametric interpretability and machine learning flexibility.

(Vallarino, 2024) A thorough and useful analysis of the machine learning environment for survival analysis is given by Vallarino (2024). The study thoroughly examines more than twenty survival models, grouping them into families like kernel-based models, multi-task learning (MTLR), tree-based techniques (Survival Trees, Random Forests), and deep learning architectures (DeepSurv, DeepHit). A key contribution of this review is a thorough examination of the benefits and drawbacks of these algorithms in relation to specific contexts, with an emphasis on how well they perform and how appropriate they are for different types of data, such as censored data, competing risks, longitudinal data, and sample size.

(ÖZCAN & KARASOY, 2019) offers a thorough analysis of the Buckley-James model as a workable substitute for conventional survival techniques. The authors describe the BJ model's theoretical underpinnings in detail, including its weight functions based on Kaplan-Meier estimates, its iterative estimation algorithm, and how regression coefficient variance is calculated. The BJ model is applied to a real-world dataset of 124 breast cancer patients and directly compared with the CPH model to illustrate its usefulness. The findings demonstrate a significant benefit of the BJ model: it does not necessitate the proportional hazards assumption, which was broken in the data on breast cancer. The literature's model selection criteria, which were based on model fit and censoring proportion, suggested that the BJ model was the better option for this dataset. The BJ model additionally offers direct estimates of expected survival times for censored individuals, and the analysis showed that radiotherapy and disease stage were significant predictors of survival time.

3. Research Methodology

3.1 Survival Analysis

The study of time-to-event data, or survival time, is the focus of a branch of statistics called survival analysis. Two examples of prospectively collected data that are commonly analyzed using survival analysis techniques are data from prospective group studies and information gathering for clinical trials. The time of origin must be identified. For instance, if the patient's survival time with a specific type of cancer is being deliberated. The time point of diagnosis for that type of cancer may be used as the time origin, in order to clearly understand the times taken into consideration, the final point or event of interest should be sufficiently described. It may then be possible to calculate the time required to travel from the time origin to the terminus. One of the primary challenges is the existence of situations where the results of an event are hard to observe after a certain amount of time (Fadhil & Al-Sarray, 2024).

3.2 Survival and Hazard Function

3.2.1 Survival Function

The likelihood that a person will live past a specific time t is known as the survival function. For instance, hours, days, months, years, cycles, and so forth are the units of t. (Taketomi et al., 2022), when $t = 0$, the survival function's initial value is 1, indicating that, survival rate of the objects observed is 100% in the beginning of the observation-that is, no events of interest have taken place. The survival function is decreasing monotonically with t (Wang et al., 2019). $S(t) = P(T > t)$ defines the survival function. Consequently, $S(t) = 1 - F(t)$, where $F(t) = 1 - S(t) = P(T \leq t)$ is the cumulative distribution function (Taketomi et al., 2022), which stands for the likelihood that the relevant event will take place before t, for continuous cases, the density function of death can be found as $f(t) = \frac{d}{dt}F(t)$; for discrete cases, it can be found as $f(t) = \frac{[F(t+\Delta t)-(F(t)]}{\Delta t}$, where Δt is a small time interval (Manzo et al., 2023).

3.2.2 Hazard Function

The hazard function $h(t)$, also known as the mortality force, the instantaneous death rate, or the conditional failure rate, is another repeatedly used function in survival analysis. The hazard function shows the event's likelihood happening at time t stated that no event has happened before time t, not the opportunity or the likelihood that the event of interest will take place. The hazard function is interpreted mathematically as follow:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr\,[t \leq T < t+\Delta t | T \geq t]}{\Delta t} = \lim_{\Delta t \to 0} \frac{F(t+\Delta t)-F(t)}{\Delta t \cdot S(t)} = \frac{f(t)}{S(t)} \quad (1)$$

Additionally, the non-negative function is $h(t)$. The hazard function can take many forms, but all survival functions $S(t)$ reduce with time. Examine the meaning of $f(t)$. The hazard function can be shown as follows since this also can be written as $f(t) = -\frac{d}{dt}S(t) \cdot \frac{1}{S(t)} = \frac{d}{dt}[lnS(t)]$ (Taketomi et al., 2022).

Therefore, the survival function was define as:

$$S(t) = P(T > t) = \exp(-G(t)) = \exp\left(-\int_0^t g(u)du\right)$$

(2)

Since $H(t)$ is the cumulative hazard function.

3.3 Censoring

The precise survival time in longitudinal studies is known only to those participants who show the relevant event during the follow-up period. All we can say about the others (those who were lost or who were disease-free at the end of the observation period) is that they did not exhibit the event of interest during the follow-up period. We refer to these people as censored observations. The ability to incorporate data from censored observation until they are eliminated from the risk set is an appealing aspect of survival analysis. When discussing censorship, the following terms are used:

- **Right censoring:** if a failure is known to occur at some point after the documented follow-up period, the subject is right censored.
- **Left censoring:** when a failure is known to have occurred prior to the documented follow-up period, this subject is left censored.
- **Interval censoring:** when an event is known to occur between two times but the precise moment of failure is unknown, the subject considered interval censored (Stevenson, 2007).

In general the censoring can be roughly divided into three categories based on the cause of the censorship types:

(a) Right censoring, in which the survival time is equal to or smaller than the observed survival time.

(b) Left censoring, a case in which the true survival time is equal to or more than the observed survival time.

(c) Interval censoring, in which we are only aware that the event takes place within a specific time frame. It supposed to be mentioned in each of the three scenarios, the actual event occurrence time is unknown (Wang et al., 2019).

3.4 Accelerated Failure Time (AFT)

The Accelerated Failure Time presuppose that the interconnection between the explanatory variables and the logarithm of survival time (T) are linear and can be received as follows:

$$\ln(T) = X\gamma + \delta\varepsilon$$

(3)

Where, the covariate matrix is denoted by X, γ represents the vector of the coefficient, δ is an unspecified scale parameter and ε is the error that go along with similar distribution to $\ln(T)$ (Abdalla & Salh, 2026); (Salih et al., 2022); (Wang et al., 2019).

3.5 Parametric Models

Many commonly used parametric distributions in survival analysis are described in this section. The scale and shape parameters are present in many distributions. Furthermore, the time t in all distributions is limited to nonnegative values $t \geq 0$ (Wang et al., 2019).

3.5.1 Exponential Distribution

With only one parameter the δ or the hazard rate parameter, the exponential model is possibly the most significant or the most straightforward of the parametric models in survival analysis. In this instance, it is assumed that the death or the failure is the independent random time to event. A peak risk or a low survival duration are indicated by a larger value of δ (Cooney & White, 2023).

3.5.2 Weibull Distribution

The Weibull model is one of the most widely used parametric distributions in survival analysis. which has two parameters $\delta > 0 \; and \; m > 0$. The shape parameter m, which provides greater flexibility than the exponential model, is used to determine the shape of the hazard function. The Weibull model will turn out to be an exponential model if $m = 1$, in which case the hazard will be a constant. The hazard function will gradually decrease if $m < 1$. The scaling parameter δ determines how the hazard function is scaled (Taketomi et al., 2022).

3.5.3 Weibull Three Parameter Distribution

The parameters for the three parameter Weibull distribution: (shape) m, (scale) ω, (location) δ. The density function (PDF) for the three parameter Weibull distribution is as follows:

$$\frac{m}{\omega}\left(\frac{t-\delta}{\omega}\right)^{m-1} e^{-\left(\frac{t-\delta}{\omega}\right)^{m}}$$

(4)

And the cumulative distribution function (CDF) is elucidated as:

$$1 - e^{-\left(\frac{t-\delta}{\omega}\right)^{m}}, \qquad t > \delta, \quad \omega > 0, m > 0.$$

(5)

The distribution profile in this model is determined by the shape parameter m. Changing m changes the distribution function's shape; for $m < 1$, the density function is shaped as J because the exponential part dominates. A skewed unimodal density curve results from the polynomial part becoming more prominent for $m > 1$. The distribution's spread is influenced by the scale parameter ω, which is measured in the same unit as t. The density is compressed or decreased when ω increases, and it is magnified or stretched when ω decreases. As a result, t varies more when ω is larger and less when ω is smaller. The distribution is shifted along the x-axis by the location parameter δ, offering a versatile model for a range of real-world situations (Salh, 2014); (Safari et al., 2025).

3.5.4 Log-normal Distribution

After exponential and Weibull distributions, the Log-normal distribution is arguably one of the most widely used distributions. For reliability analysis, such as the analysis of equipment failure time under fatigue life testing or field reliability testing, the Log-normal distribution fits data well. Its associated with normal distribution is the cause of its popularity. Nonetheless, the Log-normal distribution shows surprisingly good fits to a large number of data that have no specific connection to the normal distribution.
A random attribute t follows the Log-normal distribution if $\log(t)$ follows $N(\mu, \sigma^2)$. If t follows Log-normal distribution, then we can write $\ln(t) = \mu + \sigma V$, where V follows the standard normal $N(0,1)$ distribution (Taketomi et al., 2022).

The density, survival and hazard function for each of the (Exponential, Weibull, Weibull three parameter and Log-normal) distributions is shown in Table (1).

Table (1) Represents the Survival Function, Hazard Function and Density for (Exponential, Weibull, Weibull 3-Parameter and Log-normal) Parametric Distributions in Survival Analysis. ((Abdalla & Salh, 2026; Wang et al., 2019)

Distribution	**Survival $S(t)$**	**Hazard $h(t)$**	**Density $f(t)$**
Exponential	$e^{(-\delta t)}$	δ	$\delta\, e^{(-\delta t)}$
Weibull	$e^{(-\delta t^{m})}$	$\delta m t^{m-1}$	$\delta m t^{m-1}\, e^{(-\delta t^{m})}$
Weibull Three Parameter	$e^{-(\frac{t-\delta}{\omega})^{m}}$	$\frac{m}{\omega}(\frac{t-\delta}{\omega})^{m-1}$	$\frac{m}{\omega}(\frac{t-\delta}{\omega})^{m-1}\, e^{-(\frac{t-\delta}{\omega})^{m}}$
Log-normal	$1\text{-}\varphi(\frac{\log(t)-\vartheta}{\sigma})$	$\dfrac{\frac{1}{\sqrt{2\pi}\sigma t} e^{(-\frac{(\log(t)-\vartheta)^2}{2\sigma^2})}}{1-\varphi(\frac{\log(t)-\vartheta}{\sigma})}$	$\frac{1}{\sqrt{2\pi}\sigma t} e^{(-\frac{(\log(t)-\vartheta)^2}{2\sigma^2})}$

Source: Prepared by researchers

3.6 Robust Estimation

A robust approach, which can handle moderate departures from ideal models with a minor loss of efficiency, is frequently proposed as an alternative to the maximum likelihood. In the past few years, numerous authors have developed robust methods to deal with outliers in survival data (Bednarski and Nowak 2003; Farcomeni and Vivian 2011; Locatelli et al. 2011; Pinto et al. 2015) (Sinha, 2019). Tukey's Biweight function has been used with the following formula:

$$\beta(x, b) = \begin{cases} (1 - (\frac{x}{b})^2)^2 & \text{if } |x| < b \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

Where
b: is the constant tunning value with (b = 4.685) (Abdalla, 2016); (Abdalla & Salh, 2026); (Maronna et al., 2006).

3.7 Buckley James Estimator

The Buckley-James (BJ) method is a popular way to estimate things in the AFT model because it works well and doesn't make many assumptions about how censoring works. There are now extensions of the BJ method, including penalized versions that can handle high-dimensional censored data well. Most of the research on these extensions has been about their theoretical properties in large samples (Lee & Kim, 2025). A lot of research has been done on the Buckley–James estimator for censored data with fully observed covariates. Comparative studies have not only endorsed it as a dependable method but have also illustrated its superior efficacy in semiparametric contexts. Later theoretical research concentrated on its asymptotic characteristics, resulting in a significant discovery: The Buckley–James type estimators asymptotically correspond to a recognized category of rank-based estimators (Yu, 2011).

The Buckley James model presuppose that the time T, exhibits a linear relationship with the (x) covariate vector. Consequently, if T_i represents n uncensored lifetimes and the x_i denote the corresponding covariate vectors, we can express

$$T_i = a_0 + a'x_i + \epsilon_i, \qquad i = 1,2,3,\dots,n \tag{7}$$

Where

ϵ_i are identically independently distributed.

Since $Y_i = \min\,(T_i, D_i)$ only observed under censoring and the censoring time are D_i.

So the Buckley and James define:

$$Y_i^* = T_i\varphi_i + E(T_i|T_i > Y_i)(1-\varphi_i), \tag{8}$$

Where the censoring indicator is $\varphi_i = I(T_i \le D_i)$ (Stare et al., 2001).

3.8 Machine Learning Method

Machine learning is utilized to teach machines how to handle data more effectively. After analyzing the data. We might not be able to decipher the pattern or draw conclusions. We use machine learning in that situation. Machine learning is becoming more and more in demand due to the large number of datasets available. Machine learning is used in many sectors, including the military and healthcare, to extract pertinent data.

Arthur Samuel, a pioneer in computer game and artificial intelligence, is credited with coining the term "Machine Learning." According to him, machine learning is a "field of research that authorize computers to learn without being directly programmed." Significant advancement have been made in numerous practical disciplines during the last couple of years thanks to the benefits of machine learning techniques, including the capability to capture nonlinear relationship and the quality of the final predictions. The primary obstacle to machine learning techniques in survival analysis is the model's time estimation and the challenge of appropriately handling censored data (Wang et al., 2019).

3.8.1 Survival Trees

A collection of q nodes and their edges are called a binary tree. The sample space X is subdivided by the nodes t_l. A tree $T(\mathcal{L}) = \{1,\dots,q\}$ is created based on a learning sample $\aleph$ a splitting rule, and a stopping rule. The nodes are represented by the incidents of the elements of $T(\mathcal{L})$. A subset of the tree $\tilde{T}(\mathcal{L}) \subset T(\mathcal{L})$ represents the set of terminal nodes or leaves, or nodes that are not split. The leaves are disjoint partitions of the covariables' entire sample space ℵ: (Hothorn et al., 2004).

$$\aleph = \bigcup_{\iota \in \tilde{T}(\mathcal{L})} t_l \text{ and } t_l \cap t_m = \emptyset \text{ for all leaves } \iota \neq m \in \tilde{T}(\mathcal{L})$$
(9)

The covariate space is divided into progressively smaller areas (nodes) by survival tree algorithms that contain observations with consistent outcomes (Bertsimas et al., 2022). Regression and classification trees designed especially to deal with censored data are called survival trees. The fundamental idea behind tree models is that similar objects will be defined by the occurrence of interest so that they will be stored in the same node after data is recursively divided according to a specific splitting criterion. There are two types of splitting criteria of survival trees: those that maximize between the heterogeneity's node and those that minimize within the homogeneity's node criterion to minimize the loss function (Hothorn et al., 2004).

3.8.2 Boosting

One of the most popular ensemble techniques is the boosting algorithm, where combines base learners into a weighted sum which shows a strong learner's final output. Hothorn et al. (2006) expanded the gradient boosting algorithm to minimize the weighted risk function by iteratively fitting a set of suitably defined residuals based on the gradient descent algorithm.

$$\hat{\alpha}\tilde{U}, X = \arg min_{\alpha} \sum_{i=1}^{N} w_i \left(\tilde{U}_i - h(X_i|\alpha)\right)$$
(10)

Where:

- The pseudo response variable is $\tilde{U}$ with $\tilde{U}_i = -\frac{\partial L(y_i,\vartheta)}{\partial \vartheta} |\vartheta = \hat{f}m(X_i)$
- α is a set of parameters.
- $h(.|\alpha_{U,X})$ is the prediction by utilizing a base learner to regress U.

To optimize this problem the steps are taken as follows:

1. Initializing $\tilde{U}_i = y_i(i = 1, \ldots, n), m = 0$ and $\hat{f}0(.|\hat{\alpha}\tilde{U}, X)$; fixing the number of iterations$M(M > 1)$.
2. Fitting $h(.|\hat{\alpha}_{U,X})$ after the residual's update $\tilde{U}_i = (i = 1, \ldots, n)$.
3. Iteratively update $\hat{f}_{m+1}(.) = \hat{f}_m(.) + \gamma h(.|\hat{\alpha}_{U,X})$
   Where γ represents the step size $0 < \gamma \leq 1$.
4. Repeat the procedures in steps (2) and (3) until $m = M$ (Wang et al., 2019).

4. Application Part

4.1 A Real Dataset Study

In this paper breast cancer dataset from Kaggle platform have been used. The dataset sample size is 4024 observations including 5 attributes. Where (616) patients are dead with (15.3%) censoring rate, the response variable for each patient is the survival in months and the explanatory attributes are the (Age, Tumor Stage, Node Stage, Stage and Grade) and the Program of R Language is used in the practical part. The dataset link in Kaggle platform is (https://www.kaggle.com/code/jingrini/breast-cancer-dataset)

4.2 Discussion

In this paper several methods have been applied to breast cancer dataset and comparisons have been made to determine which parametric method performed best and to see the performance of machine learning and robust estimation method in survival analysis for the parameters:

1. The first method is buckley james estimator with parametric models such as (Exponential, Weibull, Weibull Three Parameter and Log-normal), where the best model is Log-normal parametric survival distribution since the smallest AIC and BIC were utilized respectively with (12904.10 and 12948.20) also the explanatory variables were significant to compare it with $\alpha = 0.001$.
2. The second method is the Robust parametric survival models by using Buckley James estimation method with Tukey's Biweight function and the best method among (Exponential, Weibull, Weibull Three Parameter and Log-normal) is the Robust Weibull 3-parameter parametric survival distribution by using Buckley James estimation method with Tukey's Biweight function where the smallest AIC and BIC were detected with (1532.9 and 1589.6) as well as the independent variables were significant to compare them with $\alpha = 0.001$.
3. The last method is machine learning method to build a survival tree for (Exponential, Weibull, Weibull Three Parameter and Log-normal) parametric models. The best method is the Machine learning method to build a survival tree for Log-normal distribution with the smallest AIC and BIC respectively (2099.474 and 2108.813) and the evaluation model adequacy within each terminal node (3, 4, and 5) by depending on the p-values shows that the Log-normal model is appropriate since the p-value is smaller than the $\alpha=0.001$.

Table (2): Represents the Survival rate and the Hazard rate for the Breast Cancer patients.

Sample Size	Survival Rate	Hazard Rate
4024	84.7 %	15.3 %
	Uncensored data	Censored Data
	3084	616

Source: Prepared by researchers

In table (2) the number of breast cancer patients are 4024 instances where 616 patients are censored and 3084 patients are alive.

Table (3): Represents the Classical and Robust Comparison by using AIC and BIC for (Exponential, Weibull, Weibull 3-Parameter and Log-normal) Parametric Models

MODEL PERFORMANCE							
		Parametric Models			Robust Parametric Models		
Model	Parameters	LogLik	AIC	BIC	LogLik	RAIC	RBIC
Log-normal	7	-6445.05	12904.10	12948.20	-6399.1	12812.2	12856.3
Weibull 3-Parameter	9	-7094.59	14207.19	14263.89	-757.5	1532.9	1589.6
Weibull	8	-7101.32	14218.65	14269.05	-6247.8	12511.6	12562.0
Exponential	7	-12956.88	25927.76	25971.86	-12059.2	24133.2	24177.3

Source: Prepared by researchers

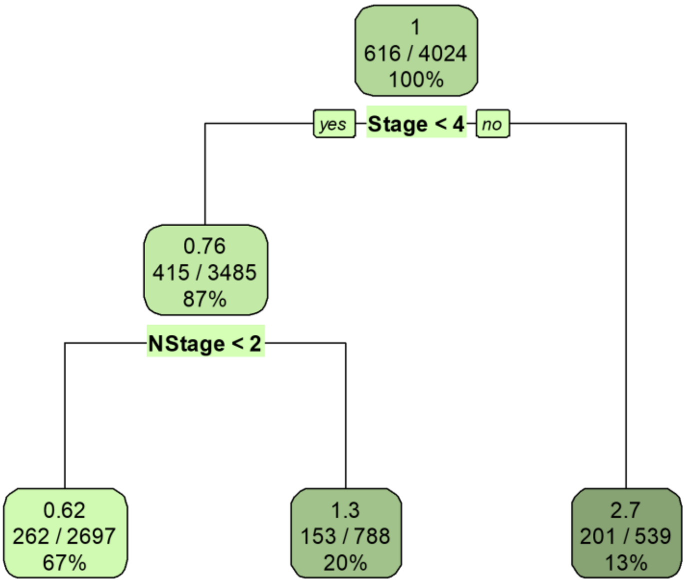
Table (3) shows the Akaike information criterion and Bayesian information criterion for both of classical and robust parametric models where, in the parametric models Log-normal distribution has the smallest AIC and BIC, while in the robust parametric models the three parameter Weibull distribution has the lowest AIC and BIC.

Table (4): Represents the Information Criteria (AIC and BIC) for the (Exponential, Weibull, Weibull 3-Parameter and Log-normal) Parametric Models with Machine Learning.

Model	Node	AIC	BIC
Exponential	1	3997.512	4003.412
	2	2108.358	2113.027
	3	2472.338	2476.628
Weibull	1	38087.503	38099.302
	2	10071.017	10080.356
	3	6057.486	6066.066
Weibull 3-Parameter	1	7492.690	7510.389
	2	2833.649	2847.658
	3	2735.272	2748.141
Log-normal	1	3974.680	3986.479
	2	2099.474	2108.813
	3	2440.659	2449.238

Source: Prepared by researchers

Table (4) shows the AIC and BIC for parametric models by using machine learning method for each node for the Exponential model the smallest AIC and BIC is in node 2, Weibull distribution’s smallest AIC and BIC is in node 3, Weibull 3-Parameter distribution’s AIC and BIC is also in node 3 and Log-normal distribution fewest AIC and BIC is in node 2, among the distributions the Log-normal distribution has smallest AIC and BIC (2099.474, 2108.813).



Source: Prepared by researchers

Figure (1) Represents the Survival Tree for the Breast Cancers Patients.

The survival tree begins with all the population of the breast cancer patients where 13% of the patients are with stage CIII which in higher risk group and the 87% of the patients are in the lower risk group. Within the large group the tree separates into two groups of Node Stage which 20% are in high risks and 67% are in lower risk branches. The terminal nodes are the bottom of the survival tree where (Node 3) the low risk group has 67% associated percentage, (Node 4) the medium risk group has 20% associated percentage and (Node 5) the high risk group has 13% associated percentage.

Table (5) Represents the Buckley James Estimator for Exponential Parametric Model.

	Variables	Coefficient	S. Error	Z	P-Value	LCI.	UCI.
Exponential Parametric Model	Age	0.0582	0.0013	41.79	0.000e+00	0.055	0.061
	Tumor Stage	0.5322	0.0478	11.13	9.303e-29	0.438	0.626
	Node Stage	0.7451	0.0820	9.08	1.087e-19	0.584	0.906
	Stage	-0.5984	0.0528	-11.32	1.051e-29	-0.702	-0.494
	Grade	0.3789	0.02472	15.32	5.269e-53	0.3304	0.4274
Robust Exponential Parametric Model	Age	0.0521	0.0746	0.697	0.486	0.050	0.053
	Tumor Stage	0.6945	0.0746	9.304	1.00e-16	0.627	0.756
	Node Stage	1.0526	0.0746	10.00	1.00e-16	0.963	1.169
	Stage	-0.7259	0.0746	-9.725	1.00e-16	-0.801	-0.658
	Grade	0.4010	0.0746	5.373	7.74e-08	0.3709	0.4349

Source: Prepared by researchers

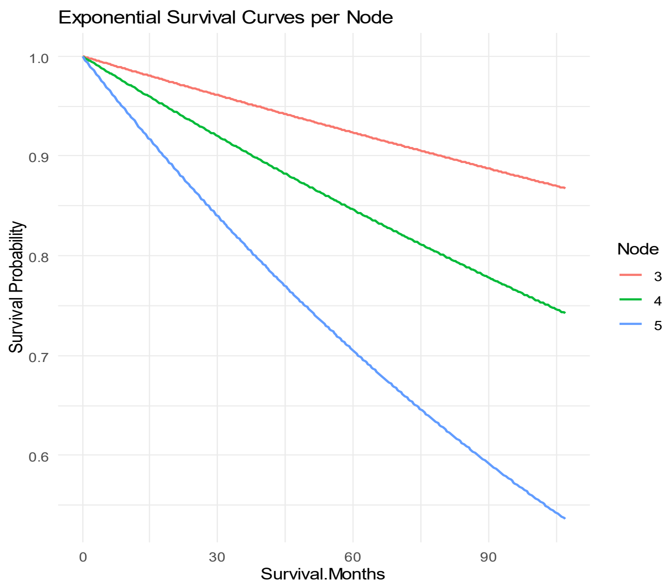
In Table (5) for both exponential parametric model and robust exponential parametric model all the independent variables are significant since the P-value is smaller than $\alpha = 0.001$, except the Age variable which is greater than $\alpha = 0.001$ by (0.486). The value of Stage in both methods are (-0.5984 and -.7259) which means it has short survival time.

Table (6) Represents the Exponential Parametric Model with Machine Learning.

Var.	Node	LogScale	SE_logscale	Z_logscale	P_logscale
1	3	6.6250	0.06626	99.972	1e-16
2	4	5.8835	0.07805	75.375	1e-16
3	5	5.1451	0.06156	83.575	1e-16

Source: Prepared by researchers

Table (6) represents the exponential parametric model were fitted separately within each terminal node, the evaluation model adequacy within each terminal node (3, 4, and 5) by depending on the p-values, it shows that the exponential model is appropriate since the p-value is smaller than the $\alpha = 0.001$.



Source: Prepared by researchers

Figure (2) Represents the Exponential Parametric Model Survival Curve.

Table (7) Represents the Buckley James Estimator for Weibull Parametric Model.

	Variables	Coefficient	SE	Z	P-Value	LCI.	UCI.
Weibull Parametric Model	Age	0.05699	0.00157	36.230	2.02e-287	0.0539	0.0600
	Tumor Stage	0.52562	0.04668	11.258	2.11e-29	0.4341	0.6171
	Node Stage	0.74125	0.07587	9.769	1.52e-22	0.5925	0.8899
	Stage	-0.58667	0.04942	-11.869	1.71e-32	-0.6835	-0.4897
	Grade	0.37597	0.02420	15.536	1.97e-54	0.32854	0.4234
Robust Weibull Parametric Model	Age	0.0441	0.0007	60.901	0.000e+00	0.042	0.046
	Tumor Stage	0.5351	0.0256	20.836	2.001e-96	0.476	0.594
	Node Stage	0.8321	0.0498	16.684	1.696e-62	0.730	0.934
	Stage	-0.5458	0.0319	-17.093	1.651e-65	-0.614	-0.478
	Grade	0.3627	0.0128	28.151	2.303e-174	0.336	0.389

Source: Prepared by researchers

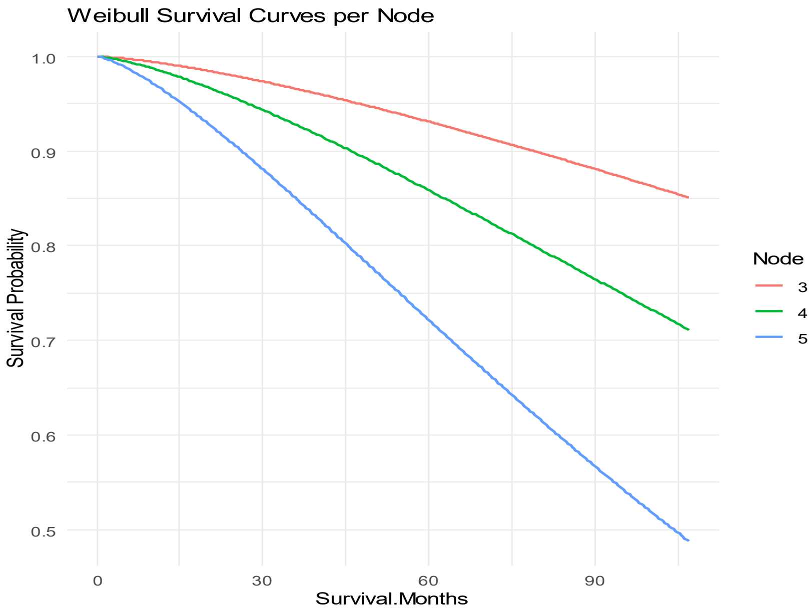
In the table above all the explanatory variables are significant in both Weibull Parametric Model and Robust Weibull Parametric Model by comparing them with $\alpha = 0.001$, and the Stage variable has the short survival time (-0.5866, -0.5458).

Table (8) Represents the Weibull Parametric Model with Machine Learning

Var.	Node	LogScale	Shape	SE_logscale	SE_Shape	Z_logscale	P_logscale
1	3	5.9640	1.4116	0.08888	0.07692	67.0978	1e-16
2	4	5.4465	1.3909	0.10158	0.10190	53.6158	1e-16
3	5	4.9156	1.3629	0.06508	0.07182	75.5319	1e-16

Source: Prepared by researchers

Table (8) represents the weibull parametric model were fitted separately within each terminal node, the evaluation model adequacy within each terminal node (3, 4, and 5) by depending on the p-values, it shows that the weibull model is appropriate since the p-value is smaller than the $\alpha = 0.001$.



Source: Prepared by researchers

Figure (3) Represents the Weibull Parametric Model Survival Curve.

Table (9) Represents the Buckley James Estimator for Three parameter Weibull Parametric Model.

	Variables	Coefficient	S. Error	Z	P-Value	LCI.	UCI.
Three parameter Weibull Parametric Model	Age	0.0571	0.0013	45.002	<0.0001	0.055	0.060
	Tumor Stage	0.5261	0.0432	12.175	<0.0001	0.441	0.611
	Node Stage	0.7424	0.0769	9.652	<0.0001	0.592	0.893
	Stage	-0.5873	0.0514	-11.428	<0.0001	-0.688	-0.487
	Grade	0.3763	0.0247	15.262	<0.0001	0.328	0.425
Robust Three parameters Weibull Parametric Model	Age	0.0467	0.0007	63.166	<0.0001	0.045	0.048
	Tumor Stage	0.5571	0.0292	19.092	<0.0001	0.500	0.614
	Node Stage	0.8503	0.0505	16.845	<0.0001	0.751	0.949
	Stage	-0.5734	0.0326	-17.615	<0.0001	-0.637	-0.510
	Grade	0.3686	0.0154	23.856	<0.0001	0.338	0.399

Source: Prepared by researchers

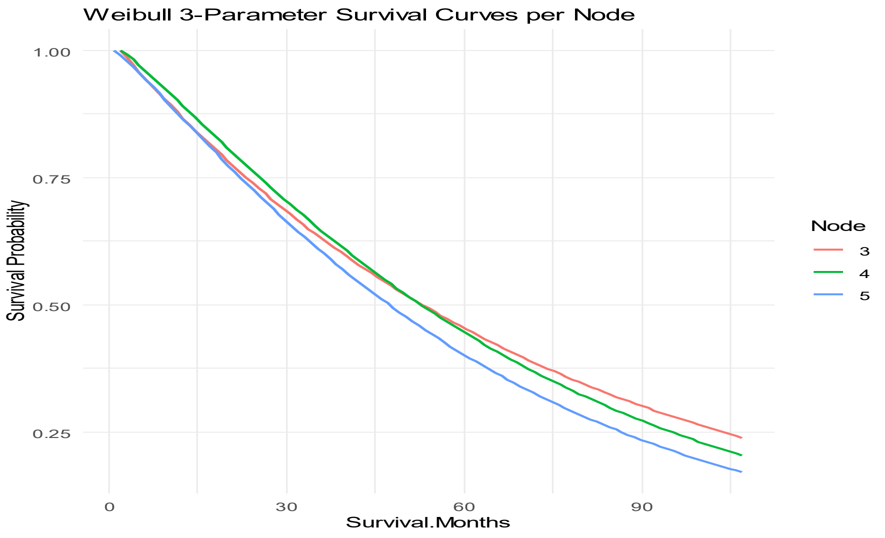
In the table above all the explanatory variables are significant in both Weibull Three Parameter Parametric Model and Robust Weibull Three Parameter Parametric Model by comparing them with $\alpha = 0.001$, and the Stage variable has the short survival time (-0.5873, -0.5734).

Table (10): Represents the Three Parameters Weibull Parametric Model with Machine Learning

Var.	Node	Shape	SE	Z	P-Value
1	3	1.0049	0.077063	13.0403	1e-16
		Scale	SE	Z	P-Value
		73.3851	23.02669	3.1869	0.00143
		Location	SE	Z	P-Value
		1.98780	0.4268	4.656	3.216e-06
2	4	Shape	SE	Z	P-Value
		1.1434	0.131193	8.71568	1e-16
		Scale	SE	Z	P-Value
		69.8743	35.00511	1.9961	0.04592
		Location	SE	Z	P-Value
		1.99736	0.7149	2.793	5.212e-03
3	5	Shape	SE	Z	P-Value
		1.1311	0.070912	15.9514	1e-16
		Scale	SE	Z	P-Value
		64.0538	21.84370	2.9323	0.00336
		Location	SE	Z	P-Value
		0.99780	1.5656	0.637	5.239e-01

Source: Prepared by researchers

Table (10) represents the Weibull 3-parameter parametric model were fitted separately within each terminal node, the evaluation model adequacy within each terminal node (3, 4, and 5) by depending on the p-values, it shows that the Weibull 3-paramter model is appropriate for node 3, since the p-value is smaller than the $\alpha = 0.001$. But Weibull 3-paramter model is not appropriate for node 4 and node 5 since the p-value is greater than the $\alpha = 0.001$.



Source: Prepared by researchers

Figure (4) Represents the Weibull 3-Parameter Parametric Model Survival Curve.

Table (11) Represents the Buckley James Estimator for Log-normal Parametric Model.

	Variables	Coefficient	S. Error	Z	P-Value	LCI.	UCI.
Log-normal Parametric Model	Age	0.06041	0.00150	40.138	0.000e+00	0.0574	0.0633
	Tumor Stage	0.56103	0.04289	13.079	4.330e-39	0.4769	0.6451
	Node Stage	0.78229	0.07995	9.785	1.310e-22	0.6255	0.9389

	Stage	-0.63121	0.05205	-12.127	7.624e-34	-0.7332	-0.5291
	Grade	0.38998	0.02385	16.351	4.267e-60	0.3432	0.4367
Robust Log-normal Parametric Model	Age	0.04409	0.000853	51.675	0.000e+00	0.042	0.046
	Tumor Stage	0.53513	0.030009	17.832	3.985e-71	0.476	0.594
	Node Stage	0.83205	0.052145	15.956	2.574e-57	0.730	0.934
	Stage	-0.54582	0.034541	-15.801	3.016e-56	-0.614	-0.478
	Grade	0.36267	0.013463	26.938	7.752e-160	0.336	0.389

Source: Prepared by researchers

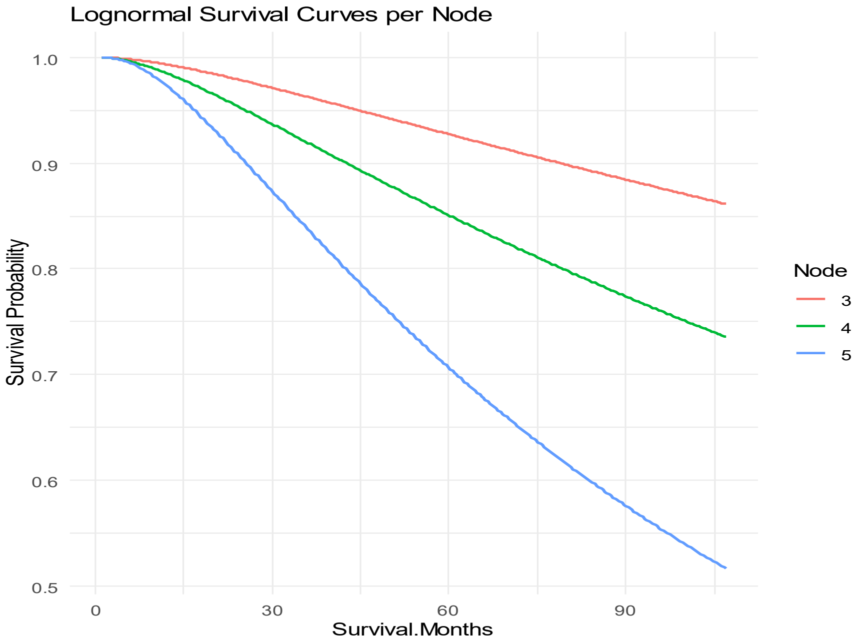
In the table above all the explanatory variables are significant in both Log-normal Parametric Model and Robust Log-normal Parametric Model by comparing them with $\alpha = 0.001$, and the Stage variable has the short survival time (-0.6312, -0.5458).

Table (12): Represents the Log-normal Parametric Model with Machine Learning

Var.	Node	Meanlog	Sdlog	SE_meanlog	SE_sdlog	Z_meanlog	P_meanlog
1	3	6.372756	1.564666	0.12903187	0.08358708	49.38901	1e-16
2	4	5.561588	1.412204	0.12351200	0.10603043	45.02872	1e-16
3	5	4.720102	1.153623	0.07288823	0.05795211	64.75808	1e-16

Source: Prepared by researchers

In Table (12) the Log-normal parametric model were fitted separately within each terminal node, the evaluation model adequacy within each terminal node (3, 4, and 5) by depending on the p-values, it shows that the Log-normal model is appropriate since the p-value is smaller than the $\alpha = 0.001$.



Source: Prepared by researchers

Figure (5) Represents the Log-normal Parametric Model Survival Curve.

5. Conclusion

In this study, we applied parametric survival models by using Buckley James estimation method, Robust parametric survival models by using Buckley James estimation method with Tukey's Biweight function and machine learning method to build a survival tree for

(Exponential, Weibull, Weibull Three Parameter and Log-normal) parametric models. And to spectacle the best result from these three different methods.

We conclude that:

1. Log-normal parametric survival distribution by using Buckley James estimation method is the best model since it has the smallest AIC and BIC among the other distributions.
2. Robust Weibull 3-parametr parametric survival distribution by using Buckley James estimation method with Tukey's Biweight function is the best model since it has the smallest AIC and BIC among the other distributions.
3. Machine learning method to build a survival tree for Log-normal distribution is the best model since it has the smallest AIC and BIC among the other distributions.

Acknowledgments: We thank those who helped us by providing scientific references.
Funding: No funding was provided for this research.

Author Contribution: Mrs. Hozan Taha Abdalla contributed to the research methodology and theoretical framework, while Professor Dr. Samira Muhamad Salh contributed to the practical aspects, including conclusions and recommendations, as well as conducting a comprehensive review and revisions.

Generative Artificial Intelligence and AI-Supported Technologies in the Writing Process: The researchers did not utilize artificial intelligence programs in their research.

Conflict of Interest: The authors declare that there is no conflict of interest related to the content, authorship, or publication of this research.

About Each Author:

Author 1: Mrs. Hozan Taha Abdalla holds a Master's degree in Statistics and is currently working on her PhD dissertation in Applied Statistics.

Author 2: Professor Dr. Samira Muhamad Salh holds a master's degree and a doctorate (PHD) in Statistics.

REFERENCES

1. Abdalla, H. T. (2016). Comparison between biexponential and robust biexponential nonlinear models using simulation (Master's thesis). University of Sulaimani.
2. Abdalla, H. T., & Salh, S. M. (2025). Choosing best robust estimation parameters among some exponential family distributions. Statistics, Optimization & Information Computing, 15(3), 1632–1652.
3. Bertsimas, D., Dunn, J., Gibson, E., et al. (2022). Optimal survival trees. Machine Learning, 111, 2951–3023.
4. Boček, L., & Štěpánek, L. (2025). Machine learning for survival analysis: A comparative study on intensive care unit (ICU) patient data and simulations. In Proceedings of the 20th Conference on Computer Science and Intelligence Systems (FedCSIS) (pp. 651–656).
5. Cooney, P., & White, A. (2023). Extending beyond Bagust and Beale: Fully parametric piecewise exponential models for extrapolation of survival outcomes in health technology assessment. Value in Health, 26(10), 1510–1517.

6. Fadhil, E. A., & Al-Sarray, B. (2024). Unified machine learning techniques for high-dimensional survival data analysis. Iraqi Journal of Science, 65(11), 6660–6675.
7. Hothorn, T., Lausen, B., Benner, A., & Radespiel-Tröger, M. (2004). Bagging survival trees. Statistics in Medicine, 23, 77–91
8. Kaindal, S., & Venkataramana, B. A. (2025). Comparative analysis of parametric survival models and machine learning methods in breast cancer prognosis. Scientific Reports, 15, 31288.
9. Lee, J., & Kim, J.-M. (2025). Counterfactual Q-learning via the linear Buckley–James method for longitudinal survival data. Manuscript submitted for publication.
10. Manzo, G., D'Amato, M., Mazzarella, G., Montella, M., & Arienzo, C., (2023). Breast cancer survival analysis agents for clinical decision support, Computer Methods and Programs in Biomedicine, 231, 107373.
11. Maronna, R. A., Martin, R. D., & Yohai, V. J. (2006). Robust statistics: Theory and methods. John Wiley & Sons.
12. Özcan, B., & Karasoy, D. (2019). Buckley–James model in survival analysis. Alphanumeric Journal: The Journal of Operations Research, Statistics, Econometrics and Management Information Systems, 7(2).
13. Safari, M. A. M., Masseran, N., Majid, M. H. A., et al. (2025). Robust estimation of the three-parameter Weibull distribution for addressing outliers in reliability analysis. Scientific Reports, 15, 11516.
14. Salh, S. M. (2014). Using Weibull distribution in the forecasting by applying on real data of the number of traffic accidents in Sulaimani during the period (2010-2013). *International Journal of Advancements in Research & Technology,* 3(10).
15. Salih, D. T. M., Salih, S. M., & Fage, M. M. (2022). Estimate parametric regression models depend on (time-to-event) survival time distributions with application, Al-Mustansiriyah Journal of Science, 33(4), 270–290.
16. Sinha, S. K. (2018). Robust estimation in accelerated failure time models. Lifetime Data Analysis.
17. Stare, J., Harrell, F. E., Jr., & Heinzl, H. (2000). BJ: An S-Plus program to fit linear regression models to censored data using the Buckley–James method. Computer Methods and Programs in Biomedicine, *64*(3), 199–204.
18. Stevenson, M. (2009). An introduction to survival analysis. EpiCentre, IVABS, Massey University.
19. Taketomi, N., Yamamoto, K., Chesneau, C., & Emura, T. (2022). Parametric distributions for survival and reliability analyses: A review and historical sketch. Mathematics, 10(22), 3907.
20. Vallarino, D. (2024). Machine learning algorithms for survival analysis: Advantages, disadvantages, and examples. International Journal of Artificial Intelligence & Machine Learning, 4(1), 10–21.
21. Wang, P., Li, Y., & Reddy, C. K. (2019). Machine learning for survival analysis: A survey. ACM Computing Surveys, 51(6), Article 110, 1–36.
22. Yu, M. (2011). Buckley–James type estimator for censored data with covariates missing by design. Scandinavian Journal of Statistics, 38, 252–267.