IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 15, 2022, accepted April 25, 2022, date of publication May 3, 2022, date of current version May 13, 2022.

*Digital Object Identifier 10.1109/ACCESS.2022.3172348*

# An Analysis of Factors Affecting Agricultural Tractors' Reliability Using Random Survival Forests Based on Warranty Data

**ZHI-LIN ZHAO, HENG-JIE YU, AND FANG CHENG**
College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China
Corresponding author: Fang Cheng (fcheng@zju.edu.cn)

This work was supported by the National Key Research and Development Program of China under Grant 2017YFD0700205.

**ABSTRACT** Warranty data are a valuable and easily accessible data source for manufacturers to assess the reliability of products in the field. Knowledge about the relationships between products' reliability and reliability factors is beneficial for manufacturers to improve products' quality. For this motivation, based on warranty data from an agricultural machinery manufacturing company in China, random survival forests (RSF), which is a machine learning method for survival analysis and provides various interpretation tools, was applied for reliability modeling in this study. The model's performance was assessed by the Harrell's concordance index (C-index) and the integrated Brier score (IBS). Thirty-four factors from production and operation were collected. Nine most important and meaningful factors were selected to show their marginal effects and interaction effects, according to which decision rules for identifying high-risk products were extracted using classification trees. The results showed that the RSF model trained by considering the observed times as the age (C-index = 0.88, IBS = 0.089) outperformed that trained by considering the observed times as the usage (C-index = 0.83, IBS = 0.15); most of the nine factors, such as "Usage Rate", had nonlinear impacts on the reliability of tractors; the marginal effects and interaction effects can be used to generate decision rules that can significantly separate high-risk products from the population. This work provides new insights for agricultural machinery manufacturers to understand their products' reliability and make reliability improvement plans and marketing plans.

**INDEX TERMS** Agricultural machinery, failure analysis, random forests, reliability engineering, warranties.

## I. INTRODUCTION

Field data about the lifetimes of products in the field are vitally important for agricultural tractor manufacturers to know the true reliability of their products. For reliability analysis, agricultural tractor manufacturers usually collect data from laboratory (reliability) tests, follow-up surveys, and onboard end devices, which are either time-consuming or expensive. Warranty data are a valuable and easily accessible data source of field data for manufacturers [1]. Compared with laboratory data, warranty data are able to capture longer time-to-failures, actual usage profiles, and the combined environmental exposures that are difficult to simulate in the laboratory [2]. A warranty is a manufacturer's assurance to a buyer that a product will perform satisfactorily over its designed useful life. Warranty data consist of claims data, which are collected during the processing of warranty claims and servicing of repairs or replacement under warranty; and supplementary data, such as manufacturing data, sales data, and maintenance data [3]. As long as manufacturers run well, warranty data can be collected without extra efforts.

One of the fundamental problems of reliability analysis is to understand the relationships between failure times and reliability factors. For reliability data, censoring occurs due to time limitations or losing track during the observation period and makes the failure times of censored products uncertain. This leads to a problem that most methods developed for normal tasks, such as classification and regression, cannot be directly applied for reliability analysis (also called survival analysis in other research fields). Traditionally, statistical approaches, such as Cox regression [4], have been widely developed to overcome the issue of censoring, but they usually have limited power in dealing with nonlinear relationships or interactions among factors. Many machine

The associate editor coordinating the review of this manuscript and approving it for publication was Yongquan Sun.

**TABLE 1. Description of the factors. The factors were sorted into production-related factors and operation-related factors.**

| Production-related Factor | Note | Operation-related Factor | Note |
|---|---|---|---|
| Model Type | Basic model type of the tractor | Usage Rate | Average usage rate of the tractor |
| Power | Rated horsepower of the model type | User Type | Type of the tractor's buyer |
| Special Model Type | Is the model type for special use | Region Sales | Province-level region cumulative sales in the year before the tractor was sold |
| Monthly Production | Number of tractors that rolled off the assembly line in the specific month | Month of Sale | Calendar month in which the tractor was sold |
| Online Time | Number of days that the tractor stayed on the assembly line | Longitude | County-level longitude of the region where the tractor was sold to |
| Month of Production | Calendar month in which the tractor rolled off the assembly line | Latitude | County-level latitude of the region where the tractor was sold to |
| C1 Supplier, C2 Supplier, …, C22 Supplier | Supplier of component C1, C2, …, C22 | | |

learning algorithms have also been adapted to tackle censored data and model covariate-dependent reliability. By transferring the original outcome to hazard ratio [5]–[7] or pseudo survival probability [8], or by using inverse probability of censoring weighting [9], general regression models, such as neural networks, can be employed for censored data. Otherwise, reducing reliability analysis task to classification task by utilizing some advanced approaches, such as multitask learning [10]–[13], is also feasible. Specifically for time series data, recurrent neural network was also extended for survival analysis. These machine learning approaches usually achieved high prediction accuracy, but their outcomes are not intuitive. Interpretation tools for them, such as survLIME and survNAM, were also developed. Reference [14] provided a thorough review of statistical methods and machine learning techniques for survival analysis where readers can get more detailed information.

RSF [15] is an ensemble method for survival analysis adapted from random forests (RF) and can deal with nonlinear relationships. RSF is non-parametric and assumption-free so that it is very practical when no clear theory or hypothesis is available for testing. More importantly, RSF provides tools for detecting important factors and interactions. RSF has been successfully applied in various fields, such as medical research [16], transportation research [17], political science [18], bankruptcy prediction [19], and assets management [20], [21].

For agricultural machinery, establishing parametric models, such as mixed Weibull distribution model, based on data from tests or surveys is the most common approach for reliability estimation. With the advancement of internet of things (IoT) technology, prognostics and health management for agricultural machinery can be achieved by monitoring the operating condition [22], such as vibration, fuel consumption, and lubricant oil consumption. However, in the field of agricultural machinery, there are few studies that have dealt with covariates in the reliability model or investigated the association between the reliability and covariates.

The objective of this study is to explore the relationships between the reliability of agricultural tractors and various factors using RSF based on warranty data. In this study, the 34 factors for the reliability of agricultural tractors were collected and were sorted into production-related factors and operation-related factors. By applying RSF on warranty data, the important factors and their interactions were detected. Furthermore, how they affected the reliability of tractors was discussed according to their marginal effects and interaction effects, based on which decision rules were also extracted.

## II. MATERIALS AND METHODS

### A. DATA PREPARATION

The warranty data used in this study were collected from a Chinese agricultural tractor manufacturing company in a period of 40 months from January 2016 through April 2019, and consisted of 44,657 tractors' information. Considering its commercially sensitive nature, the data were masked to protect proprietary information. The data were imported from the company's manufacturing execution system and enterprise resource planning system, and were aggregated into 34 factors. Table 1 lists the 34 factors and their descriptions.

The factors were sorted into production-related and operation-related factors so that the results can give feedback to the production and sales process. Production-related factors consist of factors that reflect products' attributes, e.g., "Model Type", "Power" and "Special Model Type"; factors that reflect assembly process, e.g., "Monthly Production", "Online Time", "Month of Production"; and factors that reflect component supplier, e.g., "C1 Supplier". Operation-related factors consist of factors that reflect working context, e.g., "Longitude". Most of the factors were original data recorded by the systems, whereas some factors were calculated, namely the "Online Time", "Usage Rate", and "Region Sales". Table 2 shows the descriptive statistics of the factors.

For a given product $i$, the necessary data can be represented by a triplet $(X_i, t_i, \delta_i)$, where $X_i$ is the factor vector; $t_i$ denotes the observed time; $\delta_i$ is the binary failure-censoring indicator for $t_i$ (taking on a value of one for a failed product and zero for a censored product). Censoring occurs when either the warranty expires or data collection ends. During the data collection period, 17,766 of 44,657 tractors were reported their first failures within the two-year warranty, whereas the

**TABLE 2. Descriptive statistics of the factors. Numerical factors are presented as the range, median and interquartile range. Categorical factors are presented as the categories and frequency.**

| Factor | Range / Categories | Median and Interquartile Range / Frequency |
|---|---|---|
| **Numerical Factor** | | |
| Power | 60-240 | 120 [90; 150] |
| Monthly Production | 277-5537 | 3040 [2066; 3891] |
| Online Time | 0.08-65 | 0.91 [0.76; 1] |
| Usage Rate | 0-24 | 0.90 [0.67, 1.62] |
| Region Sales | 17-20088 | 5812 [1722, 9422] |
| Longitude | 75.87-134.31 | 115.45 [113.02, 118] |
| Latitude | 18.29-52.34 | 34.20 [32.11, 37.87] |
| **Categorical Factor** | | |
| Model Type | M1, M2, …, M63 | 661, 2358, …, 5 |
| Special Model Type | Yes=1, No=0 | 27596, 17061 |
| Month of Production | Jan, Feb, Mar, Apr, May, Jun, July, Aug, Sep, Oct, Nov, Dec | 5151, 5335, 7028, 2942, 1682, 1384, 4760, 7935, 4273, 1003, 899, 2265 |
| C1 Supplier | C1S1, C1S2, C1S3, C1S4, C1S5, C1S6 | 6900, 22843, 11597, 357, 960, 80 |
| C2 Supplier | C2S1, C2S2 | 7190, 33517 |
| C3 Supplier | C3S1,C3S2,C3S3 | 33211, 9544, 1 |
| C4 Supplier | C4S1, C4S2, C4S3, C4S4, C4S5 | 28517, 118, 9854, 2907, 398 |
| C5 Supplier | C5S1, C5S2, C5S3,C5S4, C5S5, C5S8, C5S9 | 20733, 2816, 13016, 5, 1, 31, 41, 1070, 19 |
| C6 Supplier | C6S1, C6S2, C6S3, C6S4, C6S5 | 13697, 27823, 43, 79, 5 |
| C7 Supplier | C7S1, C7S2, C7S3, C7S4, C7S5 | 30629, 7148, 3370, 1, 156 |
| C8 Supplier | C8S1, C8S2, C8S3, C8S4, C8S5, C8S6, C8S7 | 31807, 4048, 5143, 1, 519, 137, 624 |
| C9 Supplier | C9S1, C9S2, C9S3, C9S4, C9S5 | 29050, 10603, 315, 109, 2235 |
| C10 Supplier | C10S1, C10S2, C10S3, C10S4, C10S5, C10S6, C10S7 | 4241, 29660, 6540, 32, 11, 1914, 5 |
| C11 Supplier | C11S1, C11S2, C11S3, C11S4, C11S5 | 35631, 980, 2104, 2998, 386 |
| C12 Supplier | C12S1, C12S2, C12S3, C12S4, C12S5 | 4677, 7153, 25395, 4870, 1 |
| C13 Supplier | C13S1, C13S2, C13S3, C13S4, C13S5 | 35616, 977, 2116, 3006, 388 |
| C14 Supplier | C14S1, C14S2, C14S3, C14S4 | 10183, 5862, 22279, 3752 |
| C15 Supplier | C15S1, C15S2, C15S3, C15S4 | 31222, 3258, 6615, 670 |
| C16 Supplier | C16S1, C16S2, C16S3, C16S4 | 7169, 5757, 665, 13407 |
| C17 Supplier | C17S1, C17S2, C17S3, C17S4, C17S5, C17S6 | 25317, 6702, 8819, 884, 12, 31 |
| C18 Supplier | C18S1, C18S2, C18S3, C18S4, C18S5 | 32676, 6301, 475, 2373, 41 |
| C19 Supplier | C19S1, C19S2, C19S3, C19S4, C19S5 | 510, 35350, 2670, 1843, 676 |
| C20 Supplier | C20S1, C20S2, C20S3, C20S4, C20S5, C20S6 | 12353,4264,4095,10513,10865,190 |
| C21 Supplier | C21S1, C21S2, C21S3, C21S4 | 9677, 9599, 21526, 954 |
| C22 Supplier | C22S1, C22S2, C22S3 | 27876, 162, 14514 |
| User Type | Enterprise, Individual User, State Farm | 1484, 40572, 2601 |
| Month of Sale | Jan, Feb, Mar, Apr, May, Jun, July, Aug, Sep, Oct, Nov, Dec | 684, 1676, 6294, 5794, 3515, 2432, 2330, 5136, 8135, 5014, 2408, 1239 |

rest of them were right-censored (i.e., their failure times could only be asserted as greater than the observed times).

The observed time $t_i$ can be represented as age in days (calendar time) or as usage in hours (operating time). Although usage is more relevant for engineering purposes, both age and usage were considered to illustrate whether the influences of the factors keep consistent. For a product with a claim, the age is the difference between the sale date and reporting date, and the usage is the recorded operating hours on the first claim. For a product with no claim, the censored age is the difference between the sale date and the censoring date (i.e., minimum of the warranty expiration date and the end date of data collection), whereas the censored usage was unknown but can be imputed by a regression model [23].

### B. PRELIMINARIES

The reliability of a product is the probability that the product will perform its intended function for a specified time when operating under normal (or stated) environmental conditions [3]. Let $T$ be a continuous random variable denoting the time to the first failure of a product. The reliability function (also called survival function) $R(t)$ is defined to be the probability that the product survives for at least a period $t$, so that

$$R(t) = P(T > t). \tag{1}$$

The hazard function $h(t)$ describes the instantaneous failure rate,

$$h(t) = \lim_{\delta t \to 0} \frac{P(t < T \leq t + \delta t | T > t)}{\delta t}. \tag{2}$$

The cumulative hazard function (CHF) is then defined as

$$H(t) = \int_0^t h(u)du. \tag{3}$$

The Cox regression model is the most commonly used regression approach for investigating the association between the reliability and factors. It is built on the proportional hazards assumption and employs partial likelihood for the parameter estimation. The Cox regression model is given by

$$h(t, X_i) = h_0(t) \exp(\beta X_i) \tag{4}$$

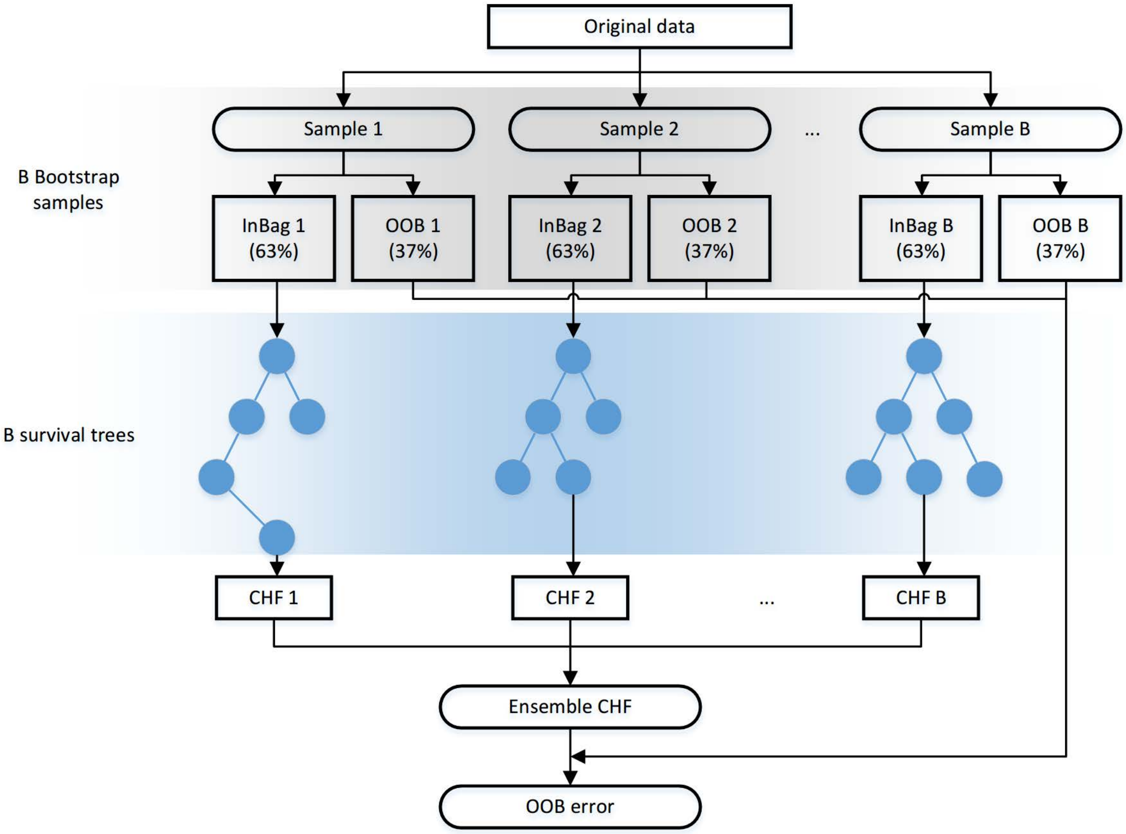


**FIGURE 1.** The flowchart of RSF adapted from [24].

where $h_0(t)$ is called the baseline hazard function and is usually not specified; $\beta$ is the coefficient vector and indicates the factors' effects. The quantity of interest from a Cox regression model is the hazard ratio (HR), which is $\exp(\beta)$. A HR smaller than one indicates reduced hazard of failure whereas a HR greater than one indicates an increased hazard of failure.

### C. RANDOM SURVIVAL FORESTS

RSF is a modification of RF for right-censored data and survival analysis. Same as in RF, randomization is introduced in RSF in two forms: growing trees using bootstrap samples and splitting trees using randomly selected subsets of features (factors). Randomization and ensemble enable RSF to approximate rich classes of functions while maintaining low generalization error. To extend RF to right-censored survival data, the following adaptions are made on RSF: splitting trees using the feature that maximizes survival difference between daughter nodes and constructing a CHF using the unique failure times at the terminal node. Fig. 1 shows the flowchart of RSF [24]. The steps for the RSF algorithm are as follows:

1) Draw $B$ bootstrap samples from the original data. Note that each bootstrap sample excludes on average 37% of the data, called out-of-bag (OOB) data.
2) Grow a survival tree for each bootstrap sample. At each node of the tree, randomly select $m$ candidate features. The node is split using the feature that maximizes survival difference between daughter nodes.
3) Grow the tree to full size under the constraint that a terminal node should have no less than $d_0 > 0$ unique events (failures).
4) Calculate a CHF for each tree. Average to obtain the ensemble CHF.
5) Using OOB data, calculate prediction error for the ensemble CHF.

In addition, RSF is capable of handling missing values. For a more in-depth explanation of this algorithm, see [15].

The RSF model was tuned on the training dataset (80% of the full dataset) using grid search and five-fold cross-validation, and then was retrained on the entire training dataset using the best set of hyperparameters. Model performance was evaluated on the test dataset (the rest 20% of the full dataset). Considering the age and the usage of tractors as the observed times, the RSF models were trained using the identical process with fixed data partition and were denoted as A-RSF and U-RSF, respectively.

Two important aspects of assessing the accuracy of the survival prediction model are discrimination and calibration. Discrimination is the ability of the model to distinguish between high- and low-risk instances, whereas calibration refers to the agreement between the observed and predicted outcomes [25]. The Harrell's concordance index

(C-index) [26] quantifies the discrimination accuracy of the survival model and indicates better accuracy with a higher value. The Brier score summarizes both calibration and discrimination prediction error simultaneously [27]. The integrated Brier score (IBS), which computes the cumulative Brier score in a specific time interval, is an overall measure for the predictive performance of the survival model at all available times and indicates better accuracy with a lower value. In this study, the predictive performance of the RSF models was presented as C-index and IBS.

## III. MODEL INTERPRETATION

When the RSF model was determined, factors with high feature importance, which optimally separate instances and contribute to the prediction, were then identified as important factors. Two separate approaches were employed to investigate the feature importance of the RSF model: permutation variable importance (VIMP), a property related to feature misspecification; and minimal depth [28], a property derived from the construction of the trees within the forest. Permutation VIMP measures the increase in the OOB error of the model after permuting the feature's OOB data, which breaks the relationship between the feature and the true outcome, and indicates a feature of predictive importance with a large value. Minimal depth measures the distance of a feature relative to the root of the tree for directly assessing the predictiveness of the feature, assuming that features with high impact on the prediction are those that most frequently split nodes nearest to the root node. Moreover, it is also possible to identify pairwise interactions among features by calculating the minimal depths of second-order maximal subtrees. A second-order maximal $(w, v)$-subtree is a maximal $w$-subtree within a maximal $v$-subtree for a feature $v$. By considering those features with closest maximal subtrees to the root node of a maximal $v$-subtree, potential interactions with $v$ can be identified.

To explore the relationships between the factors and products' reliability, the partial dependence plots (PDPs) [29] were used to show how the factors and pairwise interactions affect the RSF model's predictions. The predicted outcome of RSF can be CHF, survival function, and ensemble mortality. Ensemble mortality is defined as the expected value for the CHF summed over time, which has a natural interpretation in terms of the expected total number of failures. Thus, ensemble mortality was used as an overall indicator here representing the failure risk of products. PDP is a global explanation method for showing the marginal effect of one or two features on the predictions. The marginal effect at a particular feature value represents the average prediction if all data points are assigned that feature value [30]. An assumption of PDP is that the features were not correlated, which is naturally valid for most of the factors considered in this study due to their physical implication. Thus, PDP was qualified in this scenario.

Based on the PDPs, rules for identifying high-risk products were extracted using classification trees and were validated on the test dataset. Points in each PDP were separated into high- and low-risk groups by the median effect value and were used to train a classification tree, which was then applied to label products in the test dataset. Cox regression was then performed to estimate the HR between the high- and low-risk groups on the test dataset. The classification tree can be linearized into decision rules for identifying high-risk products if there is a significant difference between the two groups.

All analyses were completed by R software version 3.6.3 (http://www.r-project.org). The RSF method was implemented by ''randomForestSRC'' package; evaluation of the model performance was finished using ''pec'' package; the Cox regression were performed by ''survival'' package; visualization was completed using ''ggplot2'', ''ggpubr'', ''parttree'', and ''rpart.plot'' packages.

## IV. RESULTS AND DISCUSSION

After tuning and retraining the RSF model on the training dataset, A-RSF and U-RSF were determined by considering the observed times as age and usage, respectively. Both the two models used 1000 trees ($B = 1000$), 32 randomly selected features for consideration at each node split ($m = 32$), log-rank splitting rule, and the minimal terminal node size of one ($d_0 = 1$). A-RSF achieved 0.88 C-index and 0.089 IBS on the test dataset, whereas U-RSF achieved 0.83 C-index and 0.15 IBS. The result that A-RSF had better accuracy than U-RSF may result from the imputation error of censored usage.

### A. FEATURE IMPORTANCE OF THE FACTORS

Two separate measures, VIMP and minimal depth, were used to investigate the feature importance for A-RSF and U-RSF. The VIMP and minimal depth rankings of all the 34 factors for both the two models are provided in Fig. 2. The points along the red dashed line indicate where the two measures agree. Points below the red dashed line are ranked higher by VIMP than by minimal depth, indicating the factors are more sensitive to misspecification. Those above the line have a higher minimal depth ranking, indicating they are better at dividing the instances. The further the points are from the line, the more the discrepancy between the two measures.

As shown in Fig. 2, the results were somewhat different as VIMP and minimal depth use different criteria. But the feature importance for A-RSF and U-RSF were generally similar. RSF uses the mean of the minimal depth distribution as the threshold value for deciding whether a feature's minimal depth value is small enough for the feature to be classified as important. The threshold values of A-RSF and U-RSF were 14.9 and 16.8, respectively. The minimal depths of ''C2 Supplier'', which were ranked last, of A-RSF and U-RSF were 7.8 and 7.7, respectively, indicating that all 34 factors are important and should be remained. More specifically, regardless of the measure and model, the three most important production-related factors were ''Power'', ''C8 Supplier'' and ''Monthly Production''; the three most important

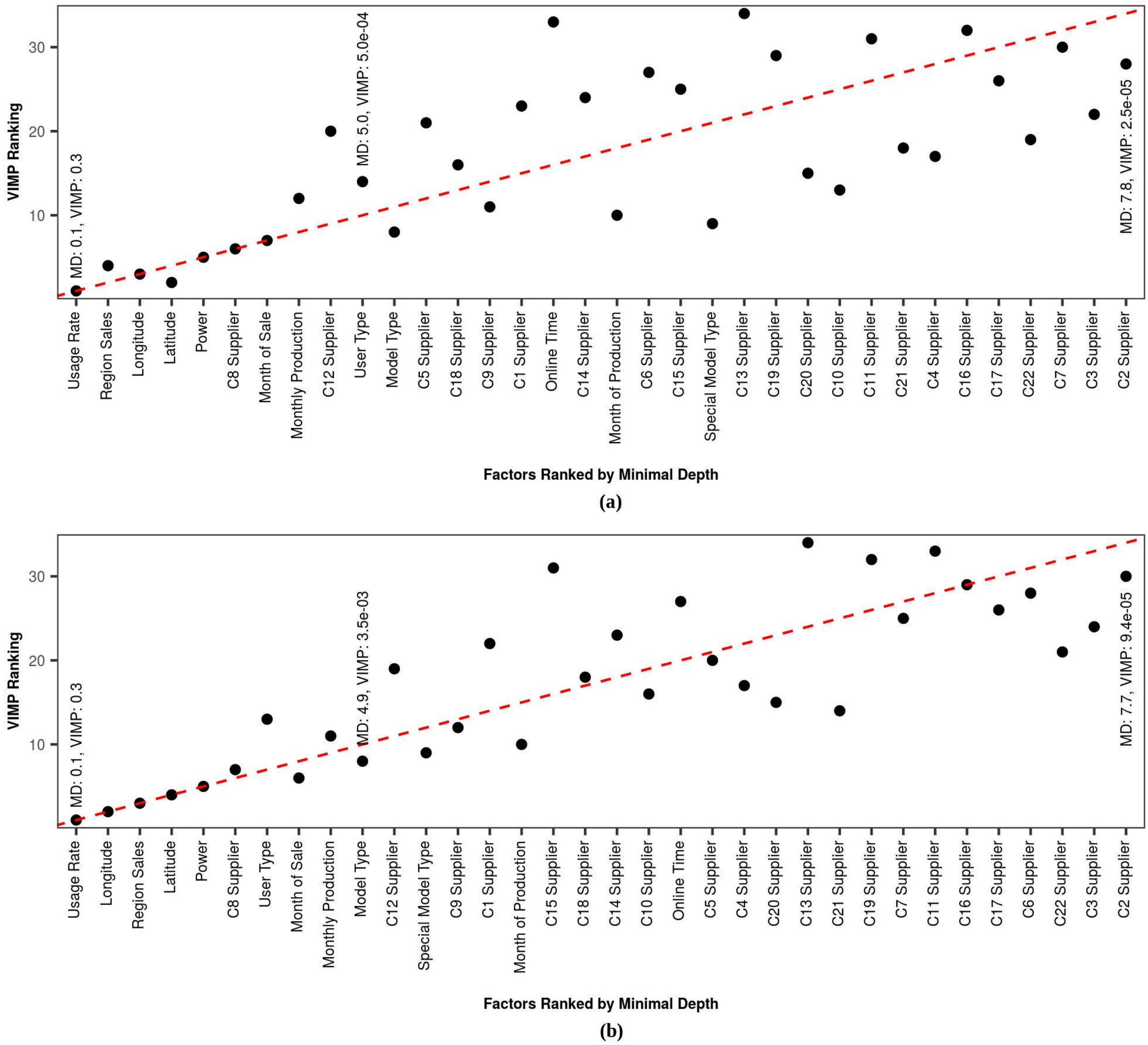


**FIGURE 2.** The rankings of VIMP and minimal depth of the 34 factors for (a) A-RSF and (b) U-RSF. Points below the red dashed line indicate factors identified as more important by VIMP than by minimal depth, and those above indicate factors identified as more important by minimal depth. The values of VIMP and minimal depth of several points are annotated.

operation-related factors were "Usage Rate", "Region Sales" and "Longitude".

For detailed illustration, in addition to the six factors mentioned above, another three less important factors, i.e., "Latitude", "Month of Sale" and "User Type", were also selected to perform in-depth analysis, considering both their feature importance and physical implication.

### *B. MARGINAL EFFECTS OF THE FACTORS*

The PDPs of the nine factors for A-RSF and U-RSF are given in Fig. 3. Mortality represents the estimated failure risk for each individual calibrated to the scale of the number of failures, and indicates a high risk with a high value. PDP shows how the average prediction of the RSF model varies over the specific feature's value. Marks (for numerical factors) and the ratio (for categorical factors) along the x-axis indicate the data distribution of each factor.

As shown in Fig. 3, the effects of the nine factors were almost consistent between A-RSF and U-RSF. The six numerical factors showed nonlinear relationships with mortality: high "Power" generated high failure risk (mortality) (Fig. 3a); either too large or too small "Monthly Production" led to high failure risk (Fig. 3b); extremely low "Usage Rate" caused abnormally high failure risk (Fig. 3d); the relationship between failure risk and "Region Sales" showed a "W" shape curve (Fig. 3e); the effects of "Longitude" (Fig. 3g) and "Latitude" (Fig. 3h) showed that tractors that worked in the central regions of China had lower failure risks.

The relationship between failure risk (mortality) and "Power" was almost monotonic, whereas other numerical factors had more complex effects. The result of "Power"

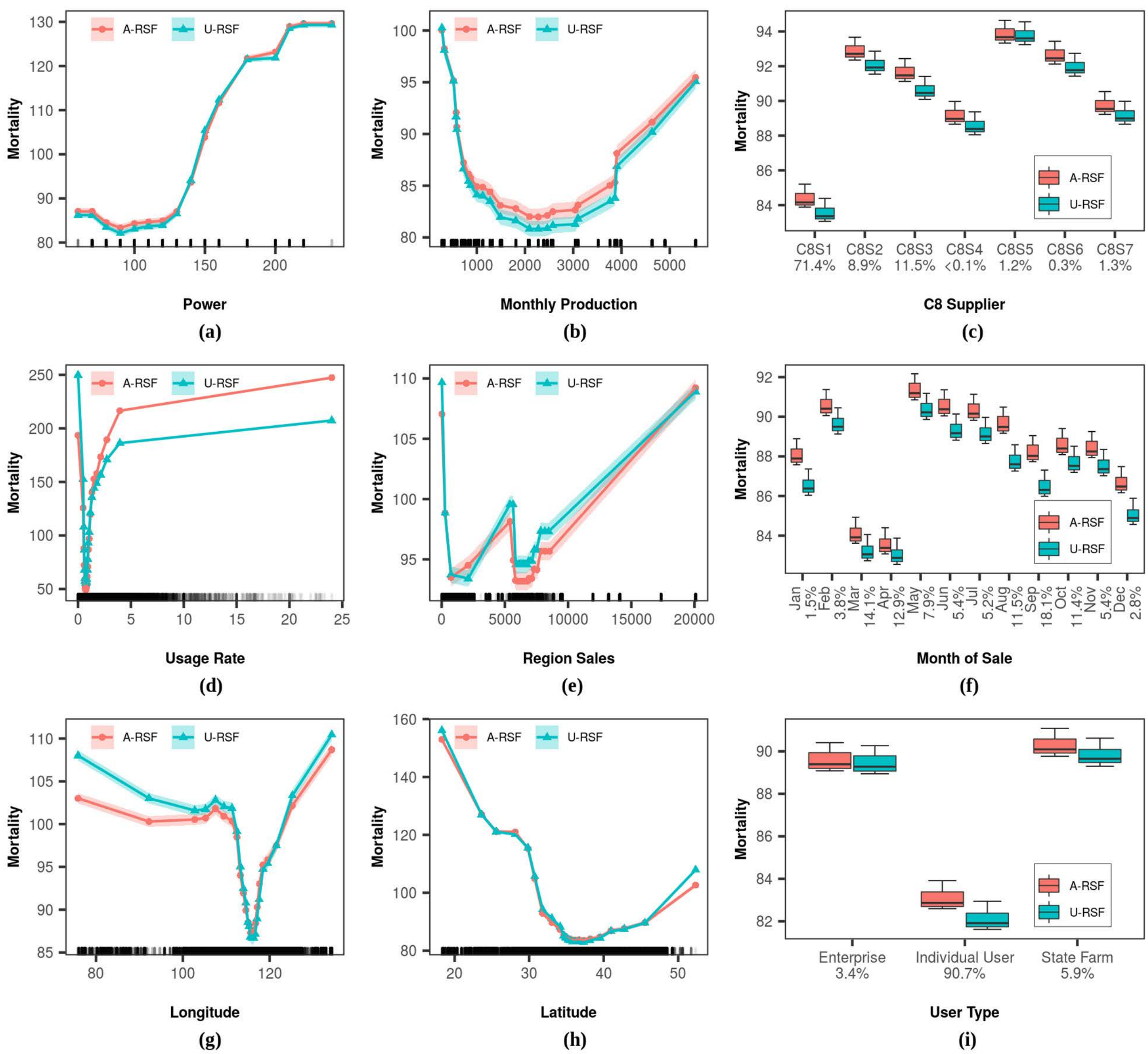


**FIGURE 3. PDPs for A-RSF and U-RSF of the nine factors: (a) "Power", (b) "Monthly Production", (c) "C8 Supplier", (d) "Usage Rate", (e) "Region Sales", (f) "Month of Sale", (g) "Longitude", (h) "Latitude", and (i) "User Type". Marks (for numerical factors) and the ratio (for categorical factors) along the x-axis indicate the data distribution of each factor. Mortality represents the average failure risk calibrated to the scale of the number of failures.**

was consistent with the company's previous surveys, in which their engineers concluded that customers of heavy-duty tractors were likely to overload tractors. The effect of "Usage Rate" was understandable (failure risk increases as usage rate increases) except for the abnormally high failure risk when "Usage Rate" was low, which may be due to the coarseness of warranty data. The effects of "Longitude" and "Latitude" may have something to do with agricultural mechanization level's region difference, as previous researches [31], [32] showed that China's central regions had a medium level of agricultural mechanization. But more dedicated works are required to validate this inference. The primary purpose of exploring the effects of "Longitude", "Latitude" and "Region sales" is to offer some insights into marketing plan making, such as what regions the company should put more maintenance resources in and how many products should be released in each region. And the effects of "Monthly Production" can also suggest a reasonable production intensity.

For categorical factors, tractors equipped with component C8 supplied by C8S1 had the highest reliability (Fig. 3c); tractors sold in March and April had lower failure risk (Fig. 3f); tractors operated by group users (i.e., enterprise and state farm) were more likely to fail (Fig. 3i). In fact, C8 refers to the tractor's front axle, which encounters the worst load conditions of the whole tractor. The result of "C8 Supplier" suggested that, for the company, the front axle's

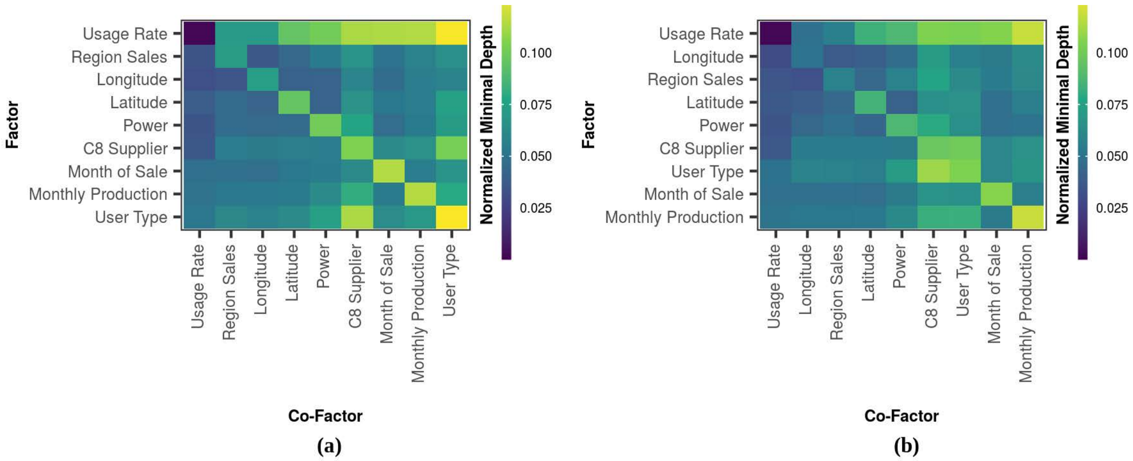


**FIGURE 4.** **Minimal depth interaction matrices of the nine factors for (a) A-RSF and (b) U-RSF. The diagonal entries are the normalized minimal depth of factor relative to the root node, whereas the off-diagonal entries indicate the normalized minimal depth of a co-factor with respect to the maximal subtree for a factor.**

quality affected the tractors' reliability most compared to other components and C8S1 should be the preferable supplier of the front axle. Group users are eager to achieve economies of scale [33]. Some possible reasoning for the effect of "User Type" is that group users may use tractors more frequently than individual users or they were more likely to report warranty claims to reduce the costs. Considering that the tractors usually failed several months after being bought, the "Month of Sale" partly reflects what seasons the tractors worked in. (The month of failure was not considered because this information of the censored tractors is unknown.) The result showed that the tractors sold in March and April had lower failure risks. Since April is the busy month for agricultural production in China, the tractors sold in March and April would go through busy months and cause high risks for the first failures. This counter-intuitive result reminds readers of the fact that all effects describe the behavior of the models and are not necessarily causal in the real world.

### C. INTERACTIONS OF THE FACTORS

Using the minimal depths of second-order maximal subtrees of the RSF model, it is also possible to calculate measures of pairwise interactions among factors. The interaction matrices of the nine factors for A-RSF and U-RSF are given in Fig. 4. The diagonal entries are the normalized minimal depth of factor relative to the root node (normalized with respect to the size of the tree), whereas the off-diagonal entries indicate the normalized minimal depth of a co-factor with respect to the maximal subtree for a factor (normalized with respect to the size of the factor's maximal subtree). Small diagonal entries indicate predictive factors. For each row, a small off-diagonal entry having a small diagonal entry is a sign of interaction between the factor and co-factor.

As shown in Fig. 4, the results were similar concerning the most important interactions for each factor. Since not all interactions with high importance were meaningful, only six pairs of factors for A-RSF were selected to show the interaction effects for illustration as given in Fig. 5. The interactions of "Usage Rate" vs. other factors were also not considered because the marginal effect of "Usage Rate" was so strong that the interaction effects always seemed less obvious.

Fig. 5 shows the two-factor PDPs for A-RSF of the six pair factors. The intersected lines indicate the presence of interaction effects between the pair of factors. Tractors that worked in the central regions of China had lower failure risks (Fig. 5a). And heavy-duty tractors that worked in low longitude regions had higher failure risks (Fig. 5b), implying that more maintenance resources of heavy-duty tractors should be placed in low longitude regions. In high longitude regions, tractors equipped with component C8 supplied by C8S3 had the highest reliability (Fig. 5c), which differs from the marginal effect of "C8 Supplier". Tractors with the rated horsepower of 160 or 180 seemed more compatible with component C8 supplied by C8S3 (Fig. 5e). The overlap between lines in Fig. 5d indicates the interaction effect of "User Type" vs. "Longitude", and the failure risks of tractors owned by different type of group users in central region had little difference. The interaction effect of "User Type" vs. "Month of Sale" was weak since there is no intersection in Fig. 5f.

### D. RULES FOR IDENTIFYING HIGH-RISK PRODUCTS

According to information on high-risk products, manufacturers can make pertinent reliability improvement plans and marketing plans. To make the above findings into decision rules for identifying high-risk products, points in each PDP were separated into high- and low-risk groups by the median mortality and were used to train a classification tree. Take the interactions of "Longitude" vs. "Latitude" as an example, the decision boundary and visualization of the classification tree are provided in Fig. 6a and Fig. 6b, respectively, which are intuitive and understandable for decision-makers to know what region they should place more maintenance resources

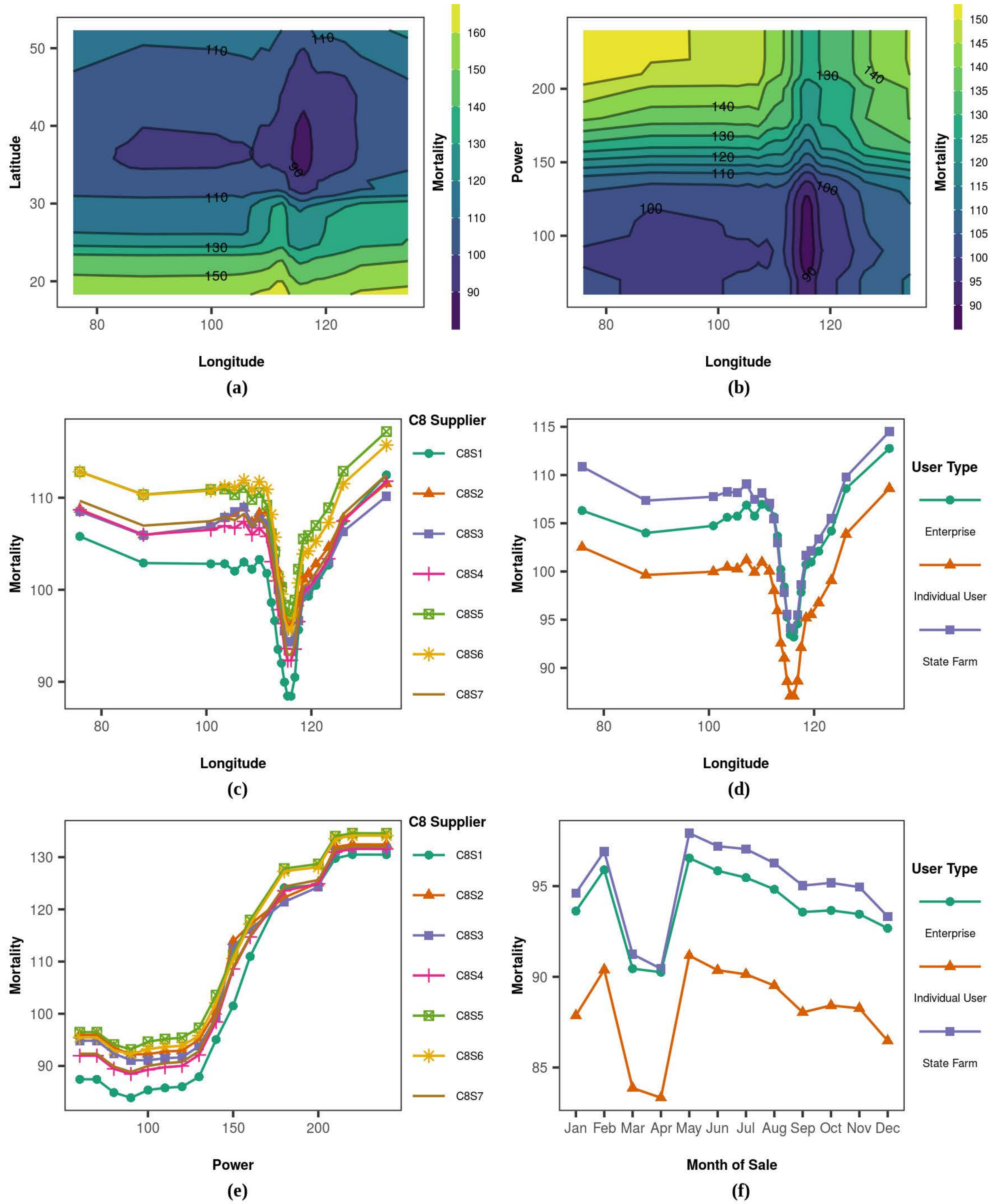


**FIGURE 5.** Two-factor PDPs for A-RSF of (a) "Longitude" vs. "Latitude", (b) "Longitude" vs. "Power", (c) "Longitude" vs. "C8 Supplier", (d) "Longitude" vs. "User Type", (e) "Power" vs. "C8 Supplier", and (f) "Month of Sale" vs. "User Type."

in according to the high failure risks. The conditions along the path in the decision tree can form a conjunction in the if-clause when the decision rule is required in text form. The classification tree was then applied to label products in the test dataset and log-rank test was performed to evaluate the survival difference between the two groups on the test dataset for validation. The log-rank test result and survival plot of products in the test dataset are given in Fig. 6c. Considering the

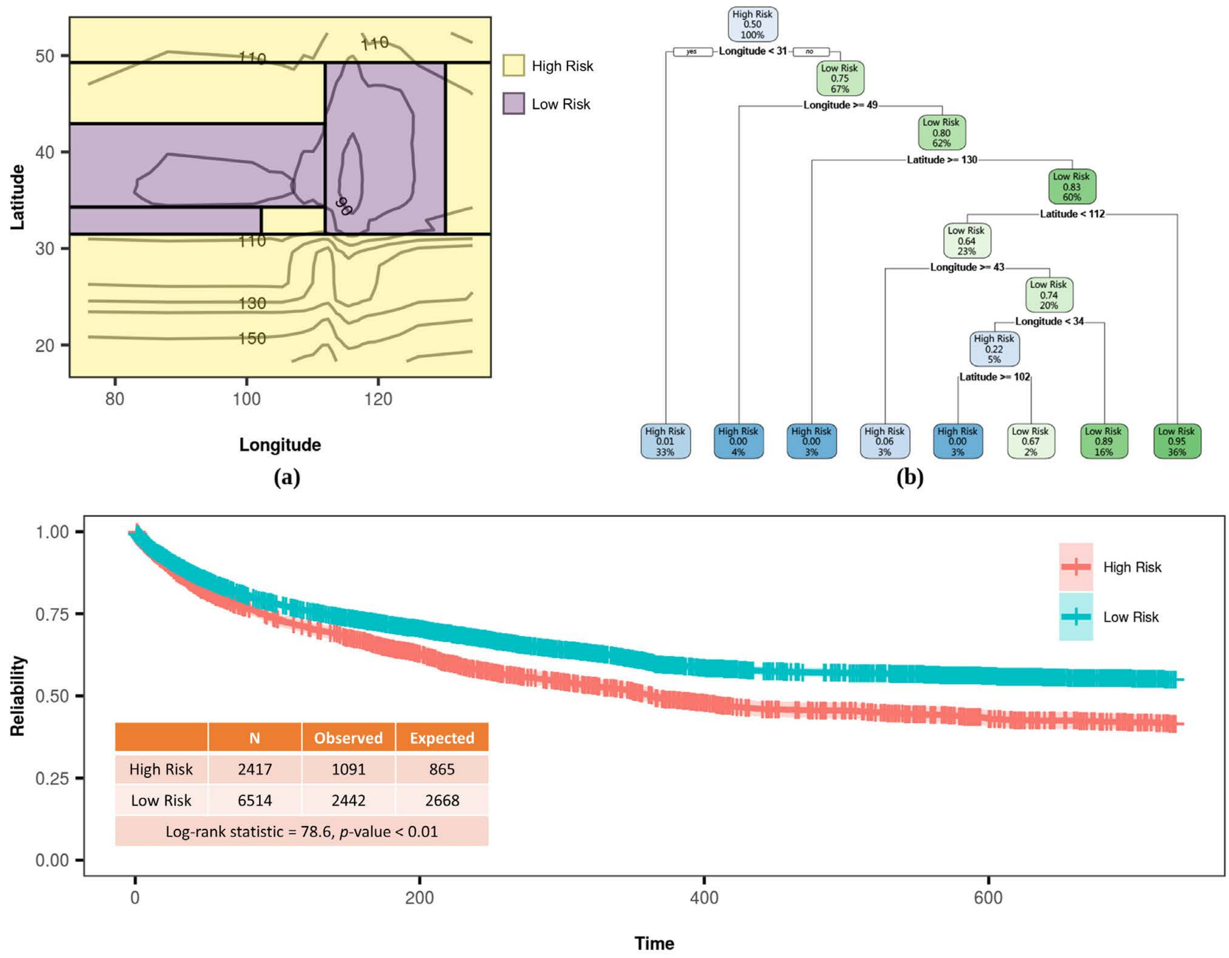


**FIGURE 6.** (a) Decision boundary and (b) visualization of the classification tree trained using the data of PDP for A-RSF of "Latitude" vs. "Longitude". (c) Survival plot and log-rank test result of high- and low-risk groups in the test dataset.

**TABLE 3.** Univariable Cox regression results of rules extracted from PDPs of A-RSF and U-RSF. Results are presented as the related factor of each rule and the corresponding hazard ratio estimated by Cox regression.

| Rule | Related Factor [a] | HR [b] A-RSF | HR [b] U-RSF |
|---|---|---|---|
| 1 | Power | 1.43*** | 1.43*** |
| 2 | Monthly Production | 0.88*** | 0.91** |
| 3 | C8 Supplier | 0.64*** | 0.64*** |
| 4 | Usage Rate | 6.12*** | 6.12*** |
| 5 | User Type | 1.21** | 1.21** |
| 6 | Region Sales | 0.99 | 0.99 |
| 7 | Month of Sale | 0.93* | 0.93* |
| 8 | Longitude | 1.23*** | 1.23*** |
| 9 | Latitude | 1.22*** | 1.21*** |
| 10 | Power vs. C8 Supplier | 0.96 | 0.96 |
| 11 | Month of Sale vs. User Type | 0.68*** | 0.68*** |
| 12 | Longitude vs. Power | 1.44*** | 1.44*** |
| 13 | Longitude vs. C8 Supplier | 1.14** | 1.14** |
| 14 | Longitude vs. User Type | 1.43*** | 1.43*** |
| 15 | Longitude vs. Latitude | 1.38*** | 1.38*** |

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05.
[a]'vs.' refers to interaction.
[b]The hazard ratio of high-risk group over low-risk group.

significance level as 5%, the significant difference between high- and low-risk groups suggested that the classification tree was a reasonable rule for identifying high-risk products.

The univariable Cox regression results of rules extracted from other PDPs are also given in Table 3. Those rules with HR greater than one, such as rules #1, can significantly identify high-risk products and keep consistent with the PDPs.

Considering the observed times as age and usage, the results of the two consideration were similar concerning feature importance, marginal effects, and interaction effects. Since A-RSF had better accuracy, age should be the preferable choice for the observed times in this scenario; however, usage might be preferable for other purposes, such as calculating mean time to failure. Noting that RSF is a data-driven method, the results can vary when the model is trained on different datasets from different scenarios; and all effects describe the behavior of the model and may not be necessarily causal in the real world.

## V. CONCLUSION

To understand the relationships between the reliability of agricultural tractors and the 34 factors from production and operation, RSF was applied on warranty data considering the observed times as age and usage. Although the factors seemed to contain more information about age, the overall results of the two considerations were similar. All 34 factors were identified as important by the RSF models. According to the marginal effects of the factors, some of the most important factors, including ''Usage Rate'', had nonlinear relationships with mortality. The interactions effects among the production-related factors and operation-related factors existed. Furthermore, decision rules, which can significantly classify tractors into high- and low-risk groups and keep consistent with the results of univariable Cox regression, can be extracted from the marginal effects and interaction effects of factors.

RSF is a promising method for analyzing factor effects on agricultural tractors' reliability, where high dimensional data and nonlinear relationships exist. Manufacturers will get insights about production and sales from the results. For further study, more information regarding production and operation, such as the primary jobs (e.g. tillage) of tractors, can be fed into the model to offer further insights. And with the development of general interpretation tools for reliability models, other appropriate machine learning methods can be performed to offer comparisons. Moreover, it is also feasible to consider the failure modes into the reliability model (i.e., multi-state model).

## REFERENCES


[1] S. Wu, ''A review on coarse warranty data and analysis,'' *Rel. Eng. Syst. Saf.*, vol. 114, pp. 1–11, Jun. 2013.

[2] S. Wu, ''Warranty data analysis: A review,'' *Qual. Rel. Eng. Int.*, vol. 28, no. 8, pp. 795–805, Dec. 2012.

[3] W. R. Blischke, M. R. Karim, and D. N. P. Murthy, *Warranty Data Collection and Analysis*. London, U.K.: Springer, 2011.

[4] D. R. Cox, ''Regression models and life-tables,'' *J. Roy. Stat. Soc. B, Methodol.*, vol. 34, no. 2, pp. 187–202, 1972.

[5] T. Ching, X. Zhu, and L. X. Garmire, ''Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data,'' *PLOS Comput. Biol.*, vol. 14, no. 4, Apr. 2018, Art. no. e1006076.

[6] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, ''DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network,'' *BMC Med. Res. Methodol.*, vol. 18, no. 1, pp. 1–12, Dec. 2018.

[7] M. Luck, T. Sylvain, H. Cardinal, A. Lodi, and Y. Bengio, ''Deep learning for patient-specific kidney graft survival analysis,'' 2017, *arXiv:1705.10245*.

[8] L. Zhao and D. Feng, ''Deep neural networks for survival analysis using pseudo values,'' *IEEE J. Biomed. Health Informat.*, vol. 24, no. 11, pp. 3308–3314, Nov. 2020.

[9] D. M. Vock, J. Wolfson, S. Bandyopadhyay, G. Adomavicius, P. E. Johnson, G. Vazquez-Benitez, and P. J. O'Connor, ''Adapting machine learning techniques to censored time-to-event health record data: A general-purpose approach using inverse probability of censoring weighting,'' *J. Biomed. Informat.*, vol. 61, pp. 119–131, Jun. 2016.

[10] S. Fotso, ''Deep neural networks for survival analysis based on a multi-task framework,'' 2018, *arXiv:1801.05512*.

[11] C. N. Yu, R. Greiner, H. C. Lin, and V. Baracos, ''Learning patient-specific cancer survival distributions as a sequence of dependent regressors,'' in *Proc. Adv. Neural Inf. Process. Syst.*, Red Hook, NY, USA, 2011, pp. 1845–1853.

[12] Y. Li, J. Wang, J. Ye, and C. K. Reddy, ''A multi-task learning formulation for survival analysis,'' in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Aug. 2016, pp. 1715–1724.

[13] C. Lee, W. R. Zame, J. Yoon, and M. Van Der Schaar, ''DeepHit: A deep learning approach to survival analysis with competing risks,'' in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018, pp. 2314–2321.

[14] P. Wang, Y. Li, and C. K. Reddy, ''Machine learning for survival analysis: A survey,'' *ACM Comput. Surv.*, vol. 51, no. 6, pp. 1–36, 2019.

[15] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S. Lauer, ''Random survival forests,'' *Ann. Appl. Statist.*, vol. 2, no. 3, pp. 841–860, Sep. 2008.

[16] Y. Yang, X. Ma, Y. Wang, and X. Ding, ''Prognosis prediction of extremity and trunk wall soft-tissue sarcomas treated with surgical resection with radiomic analysis based on random survival forest,'' *Updates Surg.*, vol. 74, no. 1, pp. 355–365, Feb. 2022.

[17] E. G. Mantouka, P. Fafoutellis, and E. I. Vlahogianni, ''Deep survival analysis of searching for on-street parking in urban areas,'' *Transp. Res. C, Emerg. Technol.*, vol. 128, Jul. 2021, Art. no. 103173.

[18] A. B. Whetten, J. R. Stevens, and D. Cann, ''The implementation of random survival forests in conflict management data: An examination of power sharing and third party mediation in post-conflict countries,'' *PLoS ONE*, vol. 16, no. 5, May 2021, Art. no. e0250963.

[19] Y. Zelenkov, ''Bankruptcy prediction using survival analysis technique,'' in *Proc. IEEE 22nd Conf. Bus. Informat.*, Antwerp, Belgium, vol. 2, Jun. 2020, pp. 141–149.

[20] D. Weeraddana, S. MallawaArachchi, T. Warnakula, Z. Li, and Y. Wang, ''Long-term pipeline failure prediction using nonparametric survival analysis,'' in *Proc. Joint. Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, Sydney, NSW, Australia, 2021, pp. 139–156.

[21] B. Snider and E. A. McBean, ''Combining machine learning and survival statistics to predict remaining service life of watermains,'' *J. Infrastruct. Syst.*, vol. 27, no. 3, Sep. 2021, Art. no. 04021019.

[22] N. Gupta, S. Gupta, M. Khosravy, N. Dey, N. Joshi, R. G. Crespo, and N. Patel, ''Economic IoT strategy: The future technology for health monitoring and diagnostic of agriculture vehicles,'' *J. Intell. Manuf.*, vol. 32, no. 4, pp. 1117–1128, Apr. 2021.

[23] Z. Zhao and F. Cheng, ''Field reliability estimation of agricultural tractors based on warranty data,'' *Trans. ASABE*, vol. 64, no. 2, pp. 705–714, 2021.

[24] V. Rodriguez-Galiano, M. P. Mendes, M. J. Garcia-Soldado, M. Chica-Olmo, and L. Ribeiro, ''Predictive modeling of groundwater nitrate pollution using random forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain),'' *Sci. Total Environ.*, vols. 476–477, pp. 189–206, Apr. 2014.

[25] M. S. Rahman, G. Ambler, B. Choodari-Oskooei, and R. Z. Omar, ''Review and evaluation of performance measures for survival prediction models in external validation settings,'' *BMC Med. Res. Methodol.*, vol. 17, no. 1, p. 60, Dec. 2017.

[26] F. E. Harrell, K. L. Lee, and D. B. Mark, ''Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors,'' *Statist. Med.*, vol. 15, no. 4, pp. 361–387, Feb. 1996.

[27] T. A. Gerds, P. K. Andersen, and M. W. Kattan, "Calibration plots for risk prediction models in the presence of competing risks," *Statist. Med.*, vol. 33, no. 18, pp. 3191–3203, Aug. 2014.
[28] H. Ishwaran, U. B. Kogalur, E. Z. Gorodeski, A. J. Minn, and M. S. Lauer, "High-dimensional variable selection for survival data," *J. Amer. Stat. Assoc.*, vol. 105, no. 489, pp. 205–217, Mar. 2010.
[29] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
[30] C. Molnar, "Global model-agnostic methods," in *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 2019. [Online]. Available: https://christophm.github.io/interpretable-ml-book/global-methods.html
[31] S. Wang, R. Zhang, D. Ai, W. Li, and W. Sun, "Measurement and spatial distribution of Chinese agricultural mechanization levels regional difference," *J. Chin. Agricult. Mech.*, vol. 37, no. 8, pp. 223–228, 2016.
[32] X. Sun, "Analysis on regional difference of agricultural machinery possession based on hierarchical clustering and GIS," *Comput. Digit. Eng.*, vol. 45, no. 6, pp. 1095–1100, 2017.
[33] C. Liu, N. Nan, and L. Huang, "Difference comparison on land intensive use of different scale farmers in rice-growing areas in Southern China," *Trans. CSAE*, vol. 34, no. 17, pp. 250–256, 2018.

**ZHI-LIN ZHAO** received the bachelor's degree in biosystem engineering from Zhejiang University, China, in 2017, where he is currently pursuing the Ph.D. degree with the College of Biosystems Engineering and Food Science. His research interests include reliability analysis of agricultural machinery and machine learning.

**HENG-JIE YU** received the B.S. degree in process equipment and control engineering from the North University of China, China, in 2019. He is currently pursuing the Ph.D. degree with the College of Biosystems Engineering and Food Science, Zhejiang University, China. His research interests include interpretable machine learning, metabolomics, and nanotechnology-enabled seed treatment.

**FANG CHENG** is currently a Professor with the College of Biosystems Engineering and Food Science, Zhejiang University, China. She has published over 80 peer-reviewed articles. Her research interests include intelligent manufacturing of agricultural machinery and nondestructive test for agricultural product.