SHAP-Driven Explainability in Survival Analysis for Predictive Maintenance Applications

Monireh Kargar-Sharif-Abad[1,*], Zahra Kharazian[1,*], Ioanna Miliou[1] and Tony Lindgren[1]

[1]*Stockholm University, Department of Computer and Systems Sciences, Kista, SE-164 07, Sweden*

Abstract

In the dynamic landscape of industrial operations, ensuring machines operate without interruption is crucial for maintaining optimal productivity levels. To avoid unexpected equipment failures, minimize downtime, and improve operational efficiency, estimating the Remaining Useful Life is very important in Predictive Maintenance. Survival analysis is a beneficial approach in this context due to its power of handling censored data (here referred to industrial assets that have not experienced a failure during the study period). Recently, with a big increase in the amount of recorded data, Machine Learning Survival models have been developed to find more complex patterns in predicting failure. However, the black-box nature of these models requires the use of explainable AI for greater transparency and interpretability. In this paper, we evaluate three Machine Learning-based Survival Analysis methods (Random Survival Forest, Gradient Boosting Survival Analysis, and Survival Support vector machine) and a traditional Survival Analysis model (Cox Proportional Hazards) using real-world data from SCANIA AB that includes 90% censored data. Results indicate that Random Survival Forest outperforms other models. In addition, we employ SHAP analysis to provide global and local explanations, highlighting the importance and interaction of features in our best-performing model. To overcome the limitation of applying SHAP on survival output, we utilize a surrogate model. Finally, SHAP identifies specific influential features, shedding light on their effects and interactions. This methodology tackles the inherent black-box nature of machine learning-based survival analysis models, providing valuable insights into their predictions. The findings from our SHAP analysis confirm the pivotal role of these identified features and their interactions, thereby enriching our comprehension of the factors influencing Remaining Useful Life predictions.

Keywords

Explainable Artificial Intelligence, Predictive Maintenance, Survival Analysis, XPdM, Censored data

1. Introduction

In the era of Industry 4.0, Predictive Maintenance (PdM) has become a cornerstone of modern manufacturing, which leverages IoT and digitization to increase machine longevity and efficiency. In fact, self-monitoring machinery that can predict and prevent failures helps minimize downtime and optimize maintenance scheduling [1]. A key aspect of PdM is to estimate indus-

HAII5.0: Embracing Human-Aware AI in Industry 5.0, at ECAI 2024, 19 October 2024, Santiago de Compostela, Spain.

*Corresponding authors.

moka6903@student.su.se (M. Kargar-Sharif-Abad); Zahra.kharazian@dsv.su.se (Z. Kharazian); ioanna.miliou@dsv.su.se (I. Miliou); tony@dsv.su.se (T. Lindgren)

https://www.su.se/profiles/zakh1874-1.623373 (Z. Kharazian); https://www.su.se/profiles/iomi2003-1.548427 (I. Miliou); https://www.dsv.su.se/~tony (T. Lindgren)

0000-0002-8430-1606 (Z. Kharazian); 0000-0002-1357-1967 (I. Miliou); 0000-0001-7713-1381 (T. Lindgren)

trial assets' Remaining Useful Life (RUL), which leads to appropriate maintenance actions, cost reductions, and operational efficiency improvements [2]. One of the significant challenges in PdM is handling censored data in which the exact failure time of components is not observed [3]. To handle censored data, the Survival analysis models (SA) were initially developed in clinical research. These models apply statistical techniques to estimate the timing of events. Traditional SA techniques, such as the Cox Proportional Hazards (CPH) model [4], provide valuable insights into survival probabilities and hazard rates [5]. However, these models have some restricted assumptions, such as linearity and proportional hazards, which limit their applicability and performance in complex industrial settings [6].

To overcome these limitations, machine learning-based survival analysis (ML-based SA) models have been developed [7, 8, 9]. These advanced models, including random survival forests and deep learning approaches, provide superior predictive performance [10] but are often criticized for their lack of transparency. The "black-box" nature of these models makes it challenging for professionals to distinguish the factors influencing the model's outputs and trust in their prediction results. Understanding these factors is essential for domain experts to trust the models and make informed maintenance decisions in real-world scenarios [11]. For instance, the feature importance analysis provides knowledge about the most influential factors that affect the failure. Moreover, it enables domain experts to estimate the usefulness of new sensors and thus enables the calculation of optimal sensor equipment.

Explainable AI (XAI) methods, such as SHAP (SHapley Additive exPlanations) [12], have been introduced to address interpretability issues in predictive models. In fact, in the field of PdM, several research studies focused on improving the interpretability of machine learning models [13, 14, 15, 16, 17]. Despite these efforts, the application of XAI to ML-based SA models in PdM remains underexplored [13]. One reason for this is that XAI methods are primarily designed for conventional machine learning models, like classifiers and regressors, which provide point predictions. In contrast, survival models produce functions, such as survival or hazard functions, representing the probability of events occurring over time rather than single-point outcomes. This difference necessitates the development of specialized XAI methods for survival models [18].

Moreover, the scarcity of real-world data in the development of RUL prediction models poses a significant challenge, especially for academia. Only a minority (24.14%) of datasets used in research accurately depict actual industrial conditions [19]. Leveraging real-world data improves the quality and cost-effectiveness of maintenance strategies and products [20].

This paper focuses on evaluating several ML-based SA models against the traditional CPH model for predicting the RUL of a specific component in truck engines manufactured by SCANIA AB in Sweden by using real-world dataset [21, 22]. The assessment is carried out by applying Harrell's concordance index (C-index), which is a standard evaluation metric in survival models, and evaluates how well they predict the ranking of survival times. Moreover, we utilize SHAP analysis to get insight into the factors that are most influential for the best-performing ML-based SA model, enhancing its interpretability. Given the challenge of applying XAI methods to SA models, we first employ a surrogate ML model to transform the SA output into a regression problem. This enables the possibility of applying XAI techniques to the surrogate regression model, thereby providing insights into the predictions of the original SA model. Overall, integrating SHAP with ML-based survival models for predictive maintenance and RUL estimation

offers a promising solution to the challenge of model interpretability, where the data include a vast majority of censored data. Consequently, this integration promotes the development of transparent and reliable predictive models that can significantly increase maintenance strategies and decision-making processes in industrial applications.

2. Related Background

2.1. Survival Analysis

Survival analysis is a statistical method that analyzes and estimates the time until an event of interest happens. More specifically, it provides insights into the probability of survival beyond a specific time point t, defined as $S(t) = \Pr(T > t)$ [3], where T represents the survival time variable. Traditional SA models vary based on their assumptions regarding the survival time distribution; non-parametric methods make no assumptions about the distribution of survival times, semi-parametric methods assume some distribution aspects, and parametric methods fully specify the distribution[3].

Survival models were initially developed in the clinical domain [23] to estimate the lifetime of patients where their end of life was not observed and was censored. Then, they expanded to other fields, such as engineering and industrial domains like PdM [5, 10]. Depending on the field of applying SA, the definition of the event of interest varies between death, recovery, failure of the machine, etc.

Censored data is inevitable and commonly encountered in real-world datasets, especially in PdM, where the quality of components is usually high. Many components may not fail within the data collection/observation time. As a consequence, the application of survival analysis models in predictive maintenance, especially in estimating RUL, is experiencing significant growth. Traditional survival analysis techniques, such as the CPH model, have been used to handle highly censored data in predicting the RUL of assets such as turbofan engines [5] and mobile working assets [24]. However, due to their strict assumptions, which may only sometimes be true [6] and the larger modeling capacity of ML models, there is increasing exploration of integrating machine learning with survival analysis.

ML-based SA models outperform traditional SA models in many cases. For instance, Voronov et al. [25] applied Random Survival Forest (RSF) to predict truck battery life, demonstrating its effectiveness with high-censoring datasets. In another study, Rahat et al. [26] found RSF superior to Gradient Boosting (GB) in predicting RUL, with a lower mean absolute error. Vallarino [6] also compared models for predicting startup failure, and the results indicated that RSF achieved the highest accuracy among the other models.

2.2. Explainability in SA

In recent years, the application of XAI in ML-based SA has attracted increasing attention from researchers across various domains [3, 27, 28]. In particular, various XAI methods have been designed and developed to interpret and explain machine learning models. Among these XAI methods, SHAP analysis is popular for interpreting ML-based SA models. This method is designed based on game theory, where every feature is considered as a game player, and

contributes a specific value to the prediction output [29, 30]. SHAP analysis interestingly provides local and global interpretability, enhancing the transparency of complex models [12]. For instance, in clinical studies, Moncada-Torres et al. [29] found that ML models, especially Extreme Gradient Boosting (XGBoost), could outperform conventional Cox models in predicting survival among breast cancer patients. They applied SHAP analysis to provide clear insights into model decisions. Moreover, Sarica et al. [30] show RSF's superior performance over Cox models, predicting Alzheimer's disease progression. Additionally, they applied SHAP on RSF output to gain more transparency on their prediction.

Integrating XAI into ML-based survival models holds significant promise for providing interpretable, accurate predictions in PdM contexts. However, further research is needed to refine these methods and fully realize their potential, particularly in the PdM domain [30, 18, 31].

3. Methodology and Problem Formulation

The overall methodology outlines four main steps: 1) Data preparation, 2) Survival modeling, 3) Regression, and 4) Explainer, as illustrated in Fig. 1 The detailed information of the steps taken in these steps are elaborated in the following section and are also summarized in Algorithm 1.

3.1. Data preparation

Let $\mathcal{D} = \{V_1, ..., V_N\}$ denote a given set of multivariate time series, V_i's, in which N is the total number of time series. In our study, the dataset is collected from three sources of information: operational, time-to-event, and specification data. For each time series V_i (here referred to as all readouts of vehicle i), the algorithm selects a random representative readout $r_{ij} \in V_i$, where j is a uniformly sampled random index, from all the readouts of vehicle i to manage the size and complexity of the dataset in processing (lines 1 and 2 of the Algorithm 1). In the next step (line 3), the algorithm converts the dataset into a format suitable for survival analysis. In this setting, each data point is characterized by three elements of (X, δ, T) where X, δ, and T are F-dimensional feature vectors, event indicators ($\delta = 1$ when experiencing the event, $\delta = 0$ in case of censoring), and observed time for individual readouts, respectively.

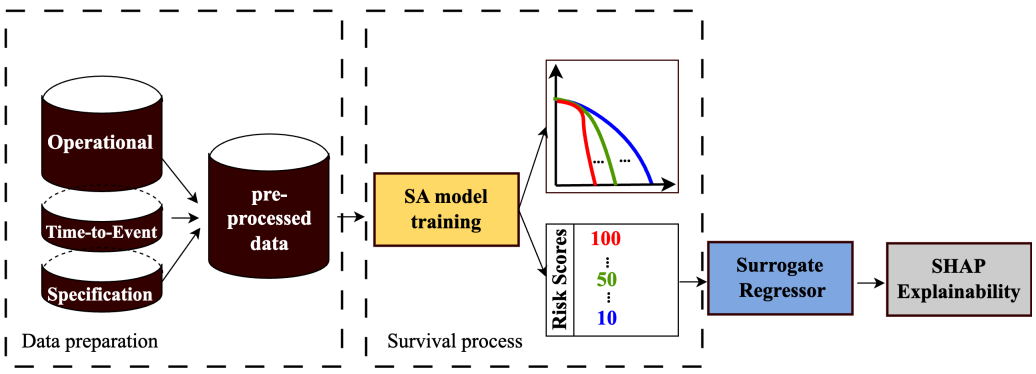


Figure 1: The methodology framework

Considering our problem, for a random readout of each vehicle, we have $r_{ij} : (X_{ij}, \delta_i, T_{ij})$, where the observed time is the true target and can be calculated for non-censored observations by $T_{ij} = T_{ij}^{\text{failure}} - T_{ij}^{\text{readout}}$. While, T_{ij} for censored observations is equal to $T_{ij} = T_{ij}^{\text{censoring}} - T_{ij}^{\text{readout}}$ where the censoring time is equal to the time of last observation. Finally, the prepared dataset undergoes the preprocessing step (line 4) to address the missing values, encode categorical features, and remove highly correlated features.

3.2. Survival process

In this step, the dataset is subsequently divided into training and testing sets. The training set is used for developing the survival models $\mathcal{M}_{\text{SA}}$ (line 6). The trained model is then employed to predict the survival curves/functions of the test samples (line 7), and these predictions are evaluated using the C-index (line 8), which is explained in Section 5.1.

3.3. Regression

To facilitate the application of explainability approaches like SHAP, a surrogate regression model $\mathcal{M}_{\text{surrogate}}$ is trained on the predictions made by the $\mathcal{M}_{\text{SA}}$ on the training data (line 9). In other words, this approach translates the complex output of the survival models (i.e., survival curves) into a compatible format (i.e., point prediction) with standard regression techniques.

3.4. Explainer

Finally, the SHAP explainer is applied to the surrogate model $\mathcal{M}_{\text{surrogate}}$ to explain the prediction output. The SHAP explainer provides a detailed understanding of how each feature contributes to the model's predictions by assigning SHAP values to each feature. For a given prediction $f(x)$ in which x is the input features, SHAP values $\phi_i(v_{f,x})$ represent the contribution of feature i. Features that provide strong power for a specific prediction will have large positive SHAP values ($\phi_i(v_{f,x}) > 0$). Conversely, uninformative features will have SHAP values close to zero ($\phi_i(v_{f,x}) \approx 0$). Features indicating a negative impact on the prediction will have negative SHAP values ($\phi_i(v_{f,x}) < 0$).

4. Empirical Evaluation

In this study, we evaluate the performance of three ML-based SA models namely RSF, Gradient Boosting Survival Analysis (GBSA), and Survival Support Vector Machines (SSVMs) against one traditional survival analysis model (i.e. CPH), using the C-index as the evaluation metric. Based on the results, the best-performing model is identified and subjected to surrogate modeling. Subsequently, SHAP analysis is employed on the output of the surrogate model, providing a comprehensive understanding of the model's predictive behavior. The SHAP analysis includes global explanations conducted for all instances in the test dataset, SHAP dependency plots applied for the four most influential features, and local explanations explored for three instances ranging from high to low risk. Through these analyses, the influence of individual features on the model's predictions is examined, highlighting both the magnitude and direction of their

Algorithm 1: Survival Analysis with SHAP Explainer for PdM

Input: Multivariate time series dataset $\mathcal{D}$, $\mathcal{M}_{\text{SA}}$, $\mathcal{M}_{\text{surrogate}}$
Output: Φ: SHAP values for explainability of the survival model

for *each time series/vehicle* $V_i \in \mathcal{D}$ **do**
    Select a random index $j \sim \text{Uniform}(1, m_i)$;
    Selected readout $r_i \leftarrow r_{ij} : (X_{ij}, \delta_i, T_{ij})$;
Preprocessing;
Data split: $\mathcal{D}$ into training set ($\mathcal{D}_{\text{train}}$) and test set ($\mathcal{D}_{\text{test}}$);
$learner = \mathcal{M}_{\text{SA}}.fit(\mathcal{D}_{\text{train}})$;
Calculate survival functions for the test set $\mathcal{D}_{\text{test}}$;
$evaluate(learner, \mathcal{D}_{\text{test}})$;
$\mathcal{M}_{\text{surrogate}}.fit(X\_train, \mathcal{M}_{\text{SA}}.predict(X\_train))$;
SHAP($\mathcal{M}_{\text{surrogate}}, \mathcal{D}_{\text{train}}$);
Compute SHAP values to explain the surrogate model's predictions;
$\Phi = \text{SHAP}(\mathcal{M}_{\text{surrogate}}, \mathcal{D}_{\text{test}})$;

impact. This approach allows the key factors affecting the model's performance and their implications for survival predictions to be elucidated. For this experiment, Python was utilized alongside packages such as scikit-survival and SHAP for model development and analysis.

4.1. Dataset

The data used in this study is a publicly available[1], real-world dataset from SCANIA AB, focusing on a specific anonymized engine component, called Component X, of SCANIA heavy trucks [21, 22]. This dataset is ideal for investigating the interpretability of survival models for predicting the RUL of heavy vehicle components, as it provides extensive real-world operational data without the need for time-consuming raw data collection.

The dataset consists of operational data, repair records (time to event), and truck specifications from 23,550 distinct trucks, organized into training, testing, and validation sets. For our analysis, we only utilized the training data because its size was sufficiently large to support our experiments. The operational dataset is a multivariate time series, with the 'time_step' column indicating the period each vehicle has been operating with Component X. It comprises 105 features which represent 14 operational variables collected by truck sensors and stored in vehicle control units. These features contain numerical data, carefully selected and anonymized by experts and named by number codes such as "Number_index". The specification dataset includes categorical data representing the truck configuration, named after seven distinct specifications and their categories [22].

The repair records dataset includes the "length_of_study_time_step" column, which shows the number of operational time steps since the component started working, and the "in_study_repair" column that serves as a class label, with 1 indicating a repair and 0 indicating no repair during the observation time. The dataset is mostly censored (90.35%), with the majority of the data

[1] https://snd.se/en/catalogue/dataset/2024-34

indicating no repair during the observation period.

4.2. Data Preprocessing

After merging the operational and specification data, the dataset contains 112 features. By enumerating the categorical data, using the get_dumies function, the number of features increases to 195. Following the removal of highly correlated features, 104 features remain. We opted to select one random readout from each vehicle. This approach also simulates real-world scenarios when the complete data is not available for each vehicle from the start of its operation. Next, the RUL is computed as explained in Section 3.1. It is worth mentioning that the dataset is split into training and testing sets, with 80% allocated for training and 20% for testing.

4.3. Survival Analysis Models

The SA models evaluated in this study include one traditional model and three ML-based models. The following sections provide a brief overview of each model.

4.3.1. Traditional Model

The CPH model is a semi-parametric survival analysis model widely employed for its interpretability and effectiveness across numerous applications [32]. It is utilized to estimate the effect of covariates on the risk of an event [4]. However, despite its popularity due to its interpretability, this model relies on strict assumptions that may not always hold true. These include assuming constant hazard ratios over time and depending on linear combinations of covariates, which may fail to capture complex and nonlinear relationships in the data.

4.3.2. ML-based Models

Random Survival Forests [9] is a variation of the traditional random forest algorithm adapted for survival analysis to predict the survival time of components. Instead of using a single decision tree, RSFs generate multiple survival trees based on bootstrapped samples of the data. The final survival prediction for a new observation is obtained by averaging the survival functions of all the trees in the forest [9].

Gradient Boosting, similar to Random Forests, is a powerful ensemble learning method known for its strong performance across various applications. In survival analysis, this technique, referred to as gradient boosting survival analysis, is adapted to predict survival times[33].

Survival Support Vector Machines are an adaptation of the traditional support vector machine framework explicitly designed for survival analysis. They handle censored data by using kernel functions, which allow for the efficient modeling of complex, high-dimensional feature spaces. This enables SSVMs to provide accurate risk predictions and effectively manage censored data in survival analysis [34].

4.3.3. Hyperparameter Tuning

The hyperparameters of the ML-based models are optimized using a grid search. A 5-fold cross-validation is performed to select hyperparameters that generalize well to unseen data.

The set of hyperparameters for each model used in grid search and the best value for each is summarized in Table 1.

Table 1
Results of hyperparameter tuning

Model	Hyperparameters	Parameter space	Best value
RSF	n_estimator	[100, 400]	100
	max_depth	[15, 20, 30]	30
	min_sample_split	[30, 40, 50]	30
	min_sample_leaf	[10, 20, 30]	20
GBSA	n_estimator	[50, 100]	50
	max_depth	[5, 6]	5
	min_sample_split	[30, 50]	50
	min_sample_leaf	[10, 20]	20
	learning_rate	[0.5, 1]	0.5
SSVM	kernel	[rbf, linear, sigmoid]	linear
	alpha	[0.0001, 3, 7, 10]	3
	gamma	[0, 0.5, 1]	0

4.4. Surrogate Model

In this study, a Random Forest (RF) regression model is employed as the surrogate function. The RF regression model is trained on the RSF output, providing point predictions, which is compatible with SHAP analysis.

4.5. Evaluation

This study uses the C-index as the assessment metric for survival analysis models due to its effectiveness in gauging predictive performance within survival analysis tasks. The C-index evaluates a model's predictive accuracy in survival analysis by assessing its ability to correctly rank pairs of samples based on their survival times. To accommodate censored observations, the index is computed by summing the concordance values for all compatible pairs and dividing by the total number of such pairs. This comprehensively evaluates the model's capability to accurately predict event order. A higher C-index denotes superior predictive accuracy, with a score of 1 signifying perfect prediction and 0.5 representing random guessing [35].

4.6. SHAP Analysis

In this study, various SHAP analysis tools are utilized to interpret the model's predictions. For global explanations, SHAP summary plots and dependency plots are used. For local explanations, SHAP force plots are employed.

5. Results

5.1. Performance evaluation

To assess the models' performance on an unseen dataset, each model was trained on the entire training dataset and then tested on the test dataset, demonstrating their generalization capabilities. Additionally, the training score for each model is reported to ensure that no overfitting has occurred. As shown in Table 2, the RSF model outperformed not only the traditional CPH but also two other ML models, namely GBSA and SSVM. This superior performance of the RSF model highlights its robustness and effectiveness in accurately predicting survival outcomes compared to the other models evaluated.

Table 2
The performance of different models

Model:	RSF	GBSA	SSVM	CPH
C-index (train)	**0.7376**	0.7297	0.6096	0.7210
C-index (test)	**0.7577**	0.7434	0.6259	0.6986

Additionally, the RSF survival probability curves for ten randomly selected instances are depicted in Figure 2, providing essential insights into the model's predictive behavior for further analysis. Lower survival probabilities (e.g., instance 460) indicate a higher risk of failure, while higher survival probabilities (e.g., instance 4699) denote healthier components with lower failure risks. Analyzing these curves allows for observation of the predicted durability and risk profiles of the selected trucks.

5.2. SHAP Analysis

The following sections discuss the results of the global explanation of the RSF prediction, as well as the local explanation for the three examples of trucks with different survival behaviors.

5.2.1. Global Explanations

SHAP summary plot Figure 3a illustrates the summary plot of SHAP values of 4710 instances (test dataset). This plot provides a global explanation of the surrogate model output across trucks of the test dataset. Each point in the plot corresponds to an individual truck. Their position along the x-axis represents the SHAP value, indicating the impact of that feature on the model's prediction for that specific truck. Features are ordered along the y-axis by their importance, determined by the mean of their absolute SHAP values. Thus, features higher on the plot are more significant to the model's overall predictions. The plot displays the SHAP values of every important feature and their impacts on the model output. The vertical axis shows the 30 most important features out of a total of 104 features, arranged in descending order of importance. Additionally, each feature is represented by a line extending from negative to positive SHAP values, color-coded with red indicating higher feature values and blue indicating

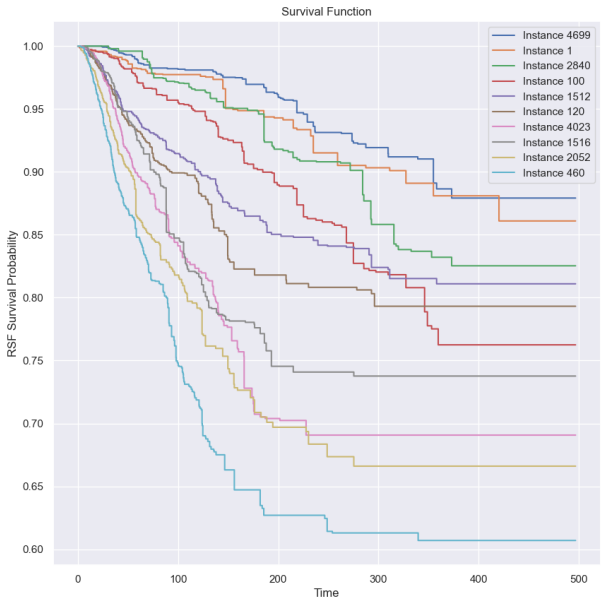
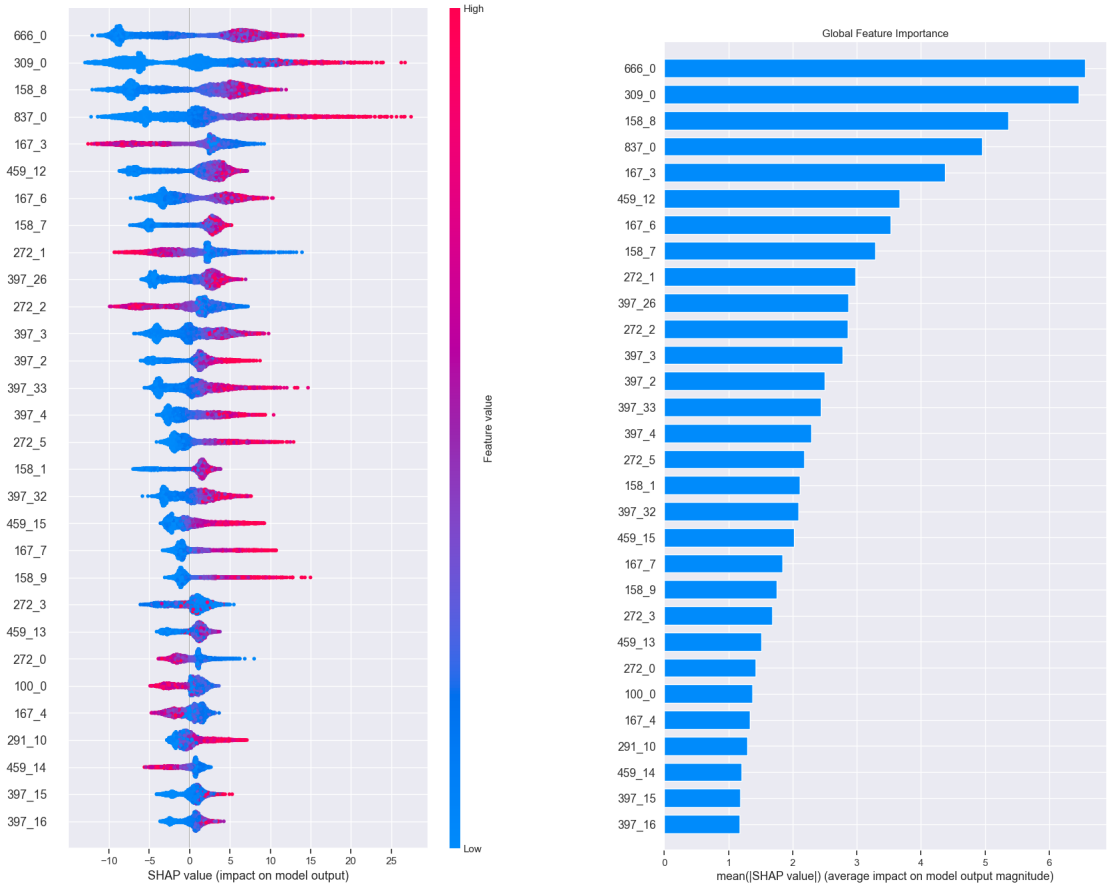


Figure 2: RSF Survival probability plot of 10 randomly selected instances

lower feature values. The negative zone represents a tendency towards censored events (y = 0), while the positive zone indicates a tendency towards failure occurrences (y = 1). In addition, the most important features not only have higher mean SHAP values but also exhibit a longer range of SHAP values along the x-axis. This indicates that these features have a more significant and varied impact on the model's predictions across different instances. As shown in Figure 3a features 666_0 and 309_0 are the most important features with the highest mean SHAP values. As the color code indicates, the higher amounts (red color) of features 666_0, 309_0, 158_8, and 837_0 and the lower amount (blue color) of feature 167_3 are associated with a greater likelihood of failure occurrence (higher SHAP value). The same interpretation applies to the rest of the features in this plot as well.

Figure 3b represents the mean SHAP value for each feature, allowing for a comparison between the features. For instance, feature 666_0 exhibits the highest impact on the model output, which is twice as influential as feature 272_2 and three times as influential as feature 158_1.

SHAP dependence plot It is a valuable tool provided by SHAP analysis. It shows how features influence the model's output and shows interactions between features. Figure 4 illustrates a SHAP force plot for the four most significant features in our study. These plots demonstrate the general influence of feature A (color-coded on the right y-axis) on feature B (x-axis for feature B's value and left y-axis for its SHAP value) across multiple instances. This plot highlights how variations in feature A affect the contribution of feature B to the model's predictions, illustrating the combined effects of these features on the overall model output. In addition, it identifies the turning points where the feature value results in a zero SHAP value, indicating a neutral impact



(a) SHAP summary plot

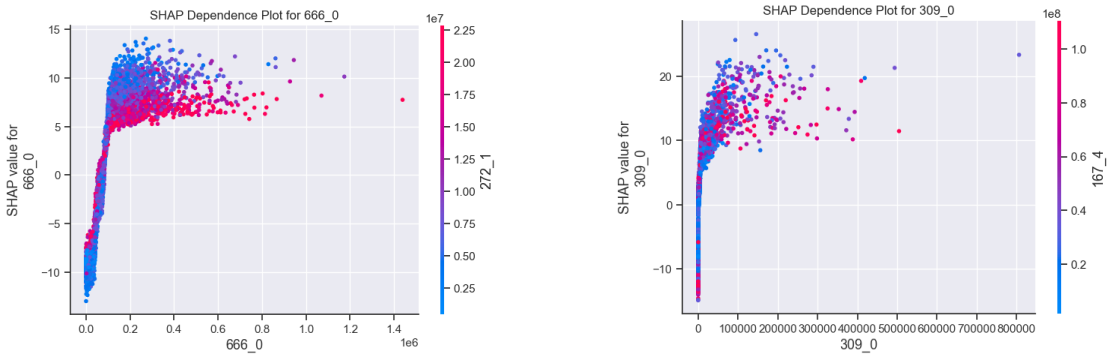
(b) Mean SHAP value

Figure 3: Global explanation of SHAP analysis on the RSF output (for all 4710 instances)

on the model's prediction at those specific values. For example, Figure 4a shows that when the feature 666_0 value surpasses 0.1e6 on the x-axis, the corresponding SHAP value turns positive. This indicates that beyond this threshold, feature 666_0 contributes to predicting failure events in the model's output. Regarding feature interaction, when feature 272_1 has a higher value (indicated by red color), the SHAP value associated with feature 666_0 moves closer to zero. This means that a high value of 272_1 reduces the impact of 666_0 on the prediction output, regardless of whether the effect of 666_0 is positive or negative. This interpretation extends to the other features shown in the Figure 4.

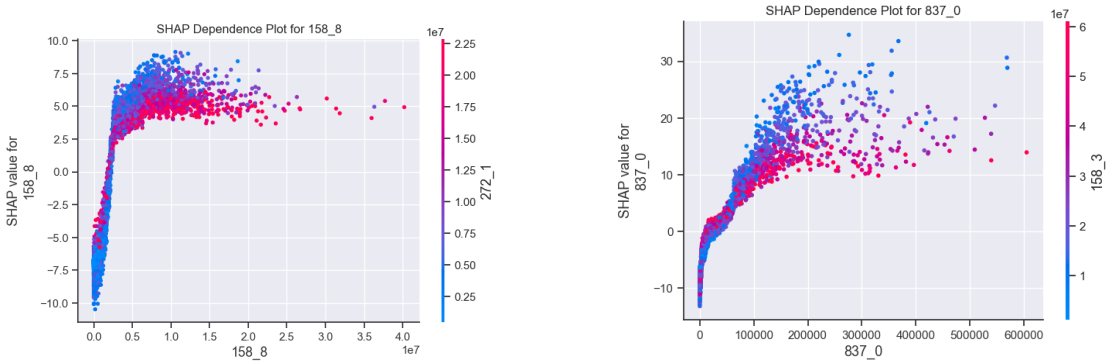
5.2.2. Local Explanations

The SHAP force diagram is a valuable tool that provides local explanations for individual predictions. It clarifies the influence of each feature on the model's prediction for a particular instance, illustrating both the direction and magnitude of each feature's impact. The blue arrows



(a) Feature 666_0

(b) Feature 309_0



(c) Feature 158_8

(d) Feature 837_0

Figure 4: SHAP dependence plots for the first four most influential features on the prediction output

pointing to the left indicate that lower values of certain features are associated with a lower model output (e.g., $y = 0$ indicating no failure occurred). Conversely, red arrows pointing to the right indicate that higher values of these features correspond to a higher model output (e.g., $y = 1$ indicating a failure occurred).

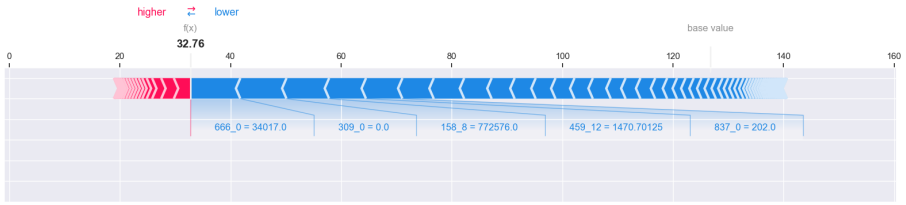


Figure 5: SHAP force plot, instance 4699, a Low-risk component.

Figures 5, 6 and 7 depict instances classified as low risk, medium risk, and high risk, respectively, based on their survival probability curves (see Figure 2). Interestingly, these instances share a common set of influential features, including 666_0, 309_0, 158_8, 837_0, and 167_3,

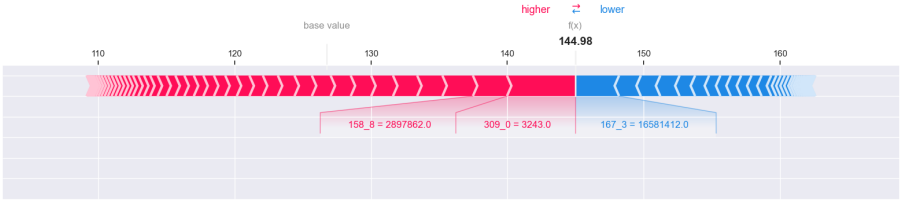


Figure 6: SHAP force plot, instance 120, a medium-risk component.

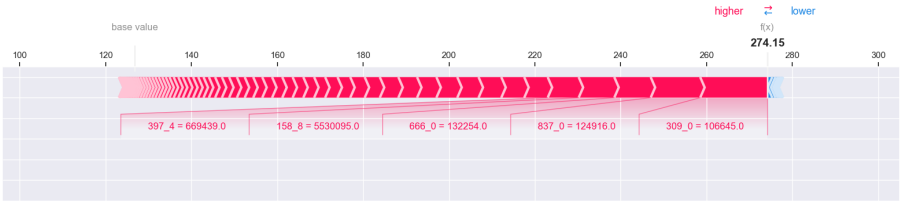


Figure 7: SHAP force plot, instance 460, a high-risk component.

as identified by SHAP global analysis (Figure 3). In addition to these common features, each instance exhibits specific features that are particularly influential. For example, in the low-risk instance (see Figure 5) feature 459_12 is important, and in the high-risk instance (see Figure 7) feature 397_4 is identified as a key feature. Addressing these features is crucial for mitigating potential failures or adverse events in maintenance scenarios.

6. Conclusion

In this paper, we tackled the challenge of incorporating explainability to survival models using real-world data, which has more than 90% censored enteries, from truck engine components manufactured by SCANIA AB. We evaluated the performance of three machine learning-based survival analysis models against the traditional Cox Proportional Hazards model for predicting the remaining useful life of truck components. The RSF model emerged as the best-performing model. To address the inherent black-box nature of ML-based survival analysis models, we utilized SHAP analysis, providing both global and local insights into feature importance and interactions. To make the RSF output compatible with SHAP analysis, first, a surrogate model was applied to the RSF output. Subsequently, SHAP analysis was exclusively applied to the surrogate model. This comprehensive approach not only identified key factors affecting model predictions but also demonstrated the potential of SHAP analysis in making complex models more transparent and understandable. Our work, in fact, can be considered as one of the first attempts to integrate XAI techniques into survival analysis, which in turn can enhance trust in predictive models and provide invaluable support for decision-making in real-world industrial scenarios.

Future work could investigate the performance of other machine learning models, particularly Deep Learning-based approaches, to enhance predictions of remaining useful life for heavy

vehicle components. Additionally, exploring changes in feature importance over time, by utilizing all operational readouts could provide insights into the dynamic nature of predictive features. Engaging domain experts, such as maintenance engineers or equipment manufacturers, in future studies, would ensure that the models incorporate relevant domain knowledge and meet practical requirements for predictive maintenance applications.

7. Acknowledgments

This work has been partially funded by Scania CV AB and the Vinnova program for Strategic Vehicle Research and Innovation (FFI) through the project RAPIDS (grant no. 2021-02522).

References

[1] Z. Li, K. Wang, Y. He, Industry 4.0-potentials for predictive maintenance, in: 6th international workshop of advanced manufacturing and automation, Atlantis Press, 2016, pp. 42–46.

[2] S. Pashami, S. Nowaczyk, Y. Fan, J. Jakubowski, N. Paiva, N. Davari, S. Bobek, S. Jamshidi, H. Sarmadi, A. Alabdallah, et al., Explainable predictive maintenance, arXiv preprint arXiv:2306.05120 (2023).

[3] P. Wang, Y. Li, C. K. Reddy, Machine learning for survival analysis: A survey, ACM Computing Surveys (CSUR) 51 (2019) 1–36.

[4] D. R. Cox, Regression models and life-tables, Journal of the Royal Statistical Society: Series B (Methodological) 34 (1972) 187–202.

[5] B. Hrnjica, S. Softic, The survival analysis for a predictive maintenance in manufacturing, in: Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems: IFIP WG 5.7 International Conference, APMS 2021, Nantes, France, September 5–9, 2021, Proceedings, Part III, Springer, 2021, pp. 78–85.

[6] D. Vallarino, Machine learning survival models restrictions: the case of startups time to failed with collinearity-related issues, Journal of Economic Statistics 1 (2023) 1–15.

[7] V. Van Belle, K. Pelckmans, S. Van Huffel, J. A. Suykens, Support vector methods for survival analysis: a comparison between ranking and regression approaches, Artificial intelligence in medicine 53 (2011) 107–118.

[8] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, Y. Kluger, Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network, BMC medical research methodology 18 (2018) 1–12.

[9] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, M. S. Lauer, Random survival forests (2008).

[10] M. Rahat, Z. Kharazian, Survloss: A new survival loss function for neural networks to process censored data, in: PHM Society European Conference, volume 8, 2024, pp. 7–7.

[11] V. Hassija, V. Chamola, A. Mahapatra, A. Singal, D. Goel, K. Huang, S. Scardapane, I. Spinelli, M. Mahmud, A. Hussain, Interpreting black-box models: a review on explainable artificial intelligence, Cognitive Computation 16 (2024) 45–74.

[12] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Advances in neural information processing systems 30 (2017).

[13] L. Cummins, A. Sommers, S. B. Ramezani, S. Mittal, J. Jabour, M. Seale, S. Rahimi, Explainable predictive maintenance: A survey of current methods, challenges and opportunities, arXiv preprint arXiv:2401.07871 (2024).
[14] S. Matzka, Explainable artificial intelligence for predictive maintenance applications, in: 2020 third international conference on artificial intelligence for industries (ai4i), IEEE, 2020, pp. 69–74.
[15] S. Vollert, M. Atzmueller, A. Theissler, Interpretable machine learning: A brief survey from the predictive maintenance perspective, in: 2021 26th IEEE international conference on emerging technologies and factory automation (ETFA), IEEE, 2021, pp. 01–08.
[16] M. Kozielski, Contextual explanations for decision support in predictive maintenance, Applied Sciences 13 (2023) 10068.
[17] C. W. Hong, C. Lee, K. Lee, M.-S. Ko, K. Hur, Explainable artificial intelligence for the remaining useful life prognosis of the turbofan engines, in: 2020 3rd ieee international conference on knowledge innovation and invention (ickii), IEEE, 2020, pp. 144–147.
[18] A. Alabdallah, S. Pashami, T. Rögnvaldsson, M. Ohlsson, Survshap: a proxy-based algorithm for explaining survival models with shap, in: 2022 IEEE 9th international conference on data science and advanced analytics (DSAA), IEEE, 2022, pp. 1–10.
[19] C. Ferreira, G. Gonçalves, Remaining useful life prediction and challenges: A literature review on the use of machine learning methods, Journal of Manufacturing Systems 63 (2022) 550–562.
[20] Y. Zhang, P. Tiňo, A. Leonardis, K. Tang, A survey on neural network interpretability, IEEE Transactions on Emerging Topics in Computational Intelligence 5 (2021) 726–742.
[21] T. Lindgren, O. Steinert, O. Andersson Reyna, Z. Kharazian, S. Magnusson, SCANIA Component X Dataset: A Real-World Multivariate Time Series Dataset for Predictive Maintenance, 2024. URL: https://doi.org/10.58141/1w9m-yz81. doi:10.58141/1w9m-yz81.
[22] Z. Kharazian, T. Lindgren, S. Magnússon, O. Steinert, O. A. Reyna, Scania component x dataset: A real-world multivariate time series dataset for predictive maintenance, arXiv preprint arXiv:2401.15199 (2024).
[23] I. Etikan, S. Abubakar, R. Alkassim, The kaplan-meier estimate in survival analysis, Biom Biostat Int J 5 (2017) 00128.
[24] Z. Yang, J. Kanniainen, T. Krogerus, F. Emmert-Streib, Prognostic modeling of predictive maintenance with survival analysis for mobile work equipment, Scientific Reports 12 (2022) 8529.
[25] S. Voronov, E. Frisk, M. Krysander, Data-driven battery lifetime prediction and confidence estimation for heavy-duty trucks, IEEE Transactions on Reliability 67 (2018) 623–639.
[26] M. Rahat, Z. Kharazian, P. S. Mashhadi, T. Rögnvaldsson, S. Choudhury, Bridging the gap: A comparative analysis of regressive remaining useful life prediction and survival analysis methods for predictive maintenance, in: PHM Society Asia-Pacific Conference, volume 4, 2023.
[27] R. Csalódi, Z. Bagyura, J. Abonyi, Mixture of survival analysis models-cluster-weighted weibull distributions, IEEE Access 9 (2021) 152288–152299.
[28] A. Kapuria, D. G. Cole, Integrating survival analysis with bayesian statistics to forecast the remaining useful life of a centrifugal pump conditional to multiple fault types, Energies 16 (2023) 3707.

[29] A. Moncada-Torres, M. C. van Maaren, M. P. Hendriks, S. Siesling, G. Geleijnse, Explainable machine learning can outperform cox regression predictions and provide insights in breast cancer survival, Scientific reports 11 (2021) 6968.
[30] A. Sarica, F. Aracri, M. G. Bianco, F. Arcuri, A. Quattrone, A. Quattrone, A. D. N. Initiative, Explainability of random survival forests in predicting conversion risk from mild cognitive impairment to alzheimer's disease, Brain Informatics 10 (2023) 31.
[31] R. Passera, S. Zompi, J. Gill, A. Busca, Explainable machine learning (xai) for survival in bone marrow transplantation trials: A technical report, BioMedInformatics 3 (2023) 752–768.
[32] M. J. Bradburn, T. G. Clark, S. B. Love, D. G. Altman, Survival analysis part ii: multivariate data analysis–an introduction to concepts and methods, British journal of cancer 89 (2003) 431–436.
[33] T. Hothorn, P. Bühlmann, S. Dudoit, A. Molinaro, M. J. Van Der Laan, Survival ensembles, Biostatistics 7 (2006) 355–373.
[34] S. Pölsterl, N. Navab, A. Katouzian, An efficient training algorithm for kernel survival support vector machines, arXiv preprint arXiv:1611.07054 (2016).
[35] I. Vasilev, M. Petrovskiy, I. Mashechkin, Sensitivity of survival analysis metrics, Mathematics 11 (2023) 4246.